

20 March 1980

Using the scientific expertise of the House of Lords

THE idea for a House of Lords select committee on science and technology came from Lords Shackleton and Sherfield last year — shortly after the reorganisation of the House of Commons select committee structure. This reorganisation has dealt a severe blow to the interests of science and science policy in Parliament because each select committee covers the remit of a single government department. Science now comes primarily under the select committee for education and science — instead of the science and technology committee of the old structure — although topics of scientific content may also touch on the interests of other committees such as agriculture or energy.

But the education and science select committee is likely to spend more time on education than scientific matters, and none of the other departments is likely to provide a proper forum for scientific discussion of issues which embrace many departments — like biotechnology, for example. Lords Sherfield and Shackleton proposed a select committee in the House of Lords to fill the gap. Parliament and government were very badly informed on science political matters, they felt, and select committees were an invaluable way of gathering information. A House of Lords committee would also have the advantage that its members could be drawn from the many respected scientists and engineers who sit in the Lords. It should therefore be able to bring more scientific expertise to bear on key issues of science policy than its predecessor in the Commons.

The first indication of its intentions came last week, when the new committee — the 'House of Lords Select Committee on Science and Technology' decided on the first two subjects for its inquiries: forestry and electric vehicles. Two subcommittees are to be set up to consider them: one, under the chairmanship of Lord Sherfield, will look at the role of fundamental research in relation to British forests and woodlands; the other, under the chairmanship of Lord Gregson, will review the case for electric vehicles in the light of present energy shortages.

These subjects are intended to fulfil the criteria laid down for topics of inquiry when the select committee was first agreed: that is they should be of interdepartmental interest, and therefore fall between the interests of several Commons select committees, or they should be of insufficient relevance to any government department to be considered at all.

Forestry is a subject which might easily have come under the Ministry of Agriculture, Fisheries and Food or the Department of the Environment. The House of Lords committee hopes to be able to ask questions which will be relevant to both departments and introduce an interdiscip-

linary approach.

Electric vehicles, on the other hand, might have come under the Department of Energy, but with the seemingly weightier problems of nuclear energy to consider, the House of Commons energy select committee would be unlikely to get round to the subject for a long time. The House of Lords committee may well, in this case, be looking at a worthy topic which might otherwise have been neglected — at least until its topicality had been lost.

Neither of these subjects, however, fires the greatest enthusiasm. They are hardly the most pressing of current issues of science and technology policy. Nevertheless they are both intended to be fairly short investigations: in particular, Lord Gregson's subcommittee on electric vehicles is due to start taking evidence shortly after Easter and to report in the autumn. Further plans are still tentative, although the committee has earmarked "aspects of information technology" and the role of science policy making bodies and their relationship to government as its next subjects for investigation. Specific topics within these broad subjects will have to be identified if useful inquiries are to be made. Under the second subject, for example, a look at the role of the National Research Development Corporation or the Department of Industry's requirements boards could be useful.

However, the committee should not shy away from considering highly political topics. Last week Lord Shackleton said that the committee had been asked "to avoid questions where the political component is much larger than the scientific". But the greater the political overtones of a scientific issue, the more important it is that it should be investigated by a select committee with scientific expertise.

Lord Shackleton also said that the committee would take an interest in "things that haven't been taken into account elsewhere". This could be incisive, if topics are chosen wisely; but it could also be a recipe for bumbledom, particularly if the committee allows itself to be steered away from the politically hot scientific topics.

One possibility that seems to have been ignored is that the committee should use its scientific expertise, and its ability to appoint scientific advisors, directly: why not look in technical depth at scientific issues which are in the province of the Commons committees, but cannot be covered there at sufficient length? Take the question raised by Sir Alan Cottrell recently about the feasibility of guarding pressurised water reactor pressure vessels against dangerous cracks. This is a fine scientific and serious political issue to which the Lords committee could pay profitable attention. □



United States

Gasohol enthusiasm may increase world food problem

A warning that growing enthusiasm for the use of agricultural crops to produce automobile fuels may contribute directly to escalating food costs and shortages has come from Lester Brown, Director of the Washington-based Worldwatch Institute.

In a report financed by the United Nations Environment Programme and the Gund Foundation, published last week, Mr Brown draws attention to potential conflicts between the different users of such crops, arguing that the demand by motorists for fuel produced in this way represents a major new variable in discussions of food supply and population growth.

"The stage is set for direct competition between the affluent minority who own the world's 315 million automobiles, and the poorest segments of humanity, for whom getting enough food to stay alive is already a struggle" Mr Brown says in his report.

The rapidly rising cost of petroleum has made the use of alcohol as a fuel — either in pure form, or mixed with petrol to produce gasohol — increasingly attractive, in addition to its environmental advantages and the fact that its production can provide an important source of jobs in developing countries.

The most advanced gasohol programme is in Brazil, where according to Mr Brown the government has set a target of producing 10.7 billion litres of alcohol by 1985. The US also has an ambitious programme, with grants and tax credits designed to reach a domestic production level of 2 billion gallons of ethanol by the mid-80s.

Plans for developing major programmes are also under way in other food exporting countries such as Australia, New Zealand

and South Africa. The consequence, says Mr Brown, is that the possibility exists for the first time of a major shift of food production capacity to non-food purposes".

Yields can be high, and could be boosted even further by a "modest effort in plant breeding and research" he suggests. As a result, the economic conditions for the large-scale commercial production from high-yielding energy crops appears to be favourable.

But the impact on food supplies and prices could be severe, particularly at a time when increases in world food output are scarcely keeping up with population growth. In Brazil, for example which already has a net deficit of 6 million metric tons a year, meeting official targets could mean planting one fifth of the country's cropland with sugarcane, driving food prices upward and "leading to more severe malnutrition among the poor".

Similarly the US, currently the world's largest supplier of grain, would have to reduce its exportable surplus considerably to meet official production goals, as distillers take a growing share of the US grain harvest. Mr Brown estimates that an individual in the US running a car solely on alcohol would require almost ten times the cropland needed to produce food for an affluent diet. Even using only a 10% ethanol mix, the car owner's cropland requirement would double.

But the solution is not to abandon gasohol. "A carefully designed alcohol fuel programme that gave farmers a first priority in the use of ethanol for tractors, farm trucks, and irrigation pumps would be a major step toward the creation of a sustainable food production system, and of a sustainable society", Mr Brown says, followed over the longer term by a programme based on forest products and cellulosic materials of forest origin. □

Sugarcane: bad-news fuel crop?



Carter: synfuels strategist

Controversy over \$20 billion synfuels programme

AFTER three months of intense negotiation, the US Congress is putting the finishing touches to legislation which will commit the government to spending \$20 billion over the next four years to stimulate the production of synthetic fuels. But the proposed legislation remains surrounded by controversy, in particular over the extent to which political interests have been allowed to override environmental and technical concerns.

The synfuels programme, a major plank in President Carter's strategy to reduce dependence on imported oil, would provide a variety of financial incentives to encourage private industry to construct plants producing synthetic oil or gas from coal, oil shale or other sources.

The proposal has been eagerly pushed by a number of constituencies, such as coal companies and engineering firms, who see it as a potential source of generous government support.

But it still faces a battery of critics, from environmentalists who argue that the health and environmental implications of a crash synfuels programme are receiving inadequate attention, to some parts of the oil industry which feel that a massive federal investment, based primarily on existing technology, could seriously distort the future growth of the industry.

So far the environmentalists have received little comfort. They have raised numerous concerns, such as the large quantities of water needed by synfuel plants, and the potential carcinogenic

'Food or fuel: new consumption for the world's cropland' Worldwatch Paper no 35 Worldwatch Institute, 1776 Massachusetts Avenue NW, Washington DC 20036. \$2.



hazards arising from the emission of aromatic and hydrocarbon amines. But legislators have indicated that they are more concerned about producing energy than meeting social and environmental costs — and are unlikely to produce much extra money for research into the latter.

Neither has much been done to calm the fears that a programme on the scale envisaged by the administration — initially aimed at producing 500,000 barrels a day by 1985 and 2 million by 1990 — is overambitious, and likely to encounter problems such as an inadequate technology base and a severe shortage of appropriately qualified scientists and engineers.

Typical reservations were expressed at a two-day meeting organised last October by the National Academy of Sciences, and attended by a cross-section of industrial and academic scientists and engineers. This reached the conclusion that the 1985 deadline could not be met, but would require six to eight years to achieve.

The meeting also pointed out that reaching the 1990 target would require second stage technologies to be selected before the results of the first stage had been properly analysed.

Following these and other similar

criticisms, the administration's initial proposals have been tempered. Thus the President has agreed that the initial cost of \$88 billion be broken down into two stages, with Congress committing itself only to \$20 billion in the first four years.

Following widespread opposition from the oil industry over what it considered to be excessive federal involvement and unfair competition, the president has also agreed that the synthetic fuels corporation which will be responsible for the programme will not have the authority to construct and run its own plants, but merely assist the private sector to do so.

Furthermore members of the Senate and the House of Representatives, meeting last week to resolve differences in their respective versions of the synfuels legislation, agreed that the deadlines for the 500,000 and 2 million barrels a day targets be set back to 1987 and 1992 respectively.

But the major thrust of the administration's strategy of pushing for ambitious production goals remains in place. These goals had been strongly supported by the Senate, but opposed by the House which had proposed a more modest package — costing only \$3 billion and with no new corporation — and

suggested a greater emphasis on conservation measures to meet the same goals.

As far as manpower needs are concerned, Dr Alex K. Logwunik of Pullman Kellogg, one of the largest engineering firms in the country, estimated last month that the proposed synfuels programme could absorb 20% to 60% of the capacity of the ten major engineering firms in the country. And he warned that the result could be a series of shortages of qualified personnel, already beginning to affect other areas of energy technology.

Supporters of the synfuels programme, however, argue that similar claims had been expressed before other large-scale national efforts, such as the Apollo moon landings. And that once the programme is under way, it will stimulate a flow of research workers and technical people from other areas.

Critics are unconvinced. But in an election year both Congress and the White House are keen to be seen to be doing something about the energy problem — and responding to the frustration of constituents takes precedence over the finer logic of technological timeliness.

David Dickson

Split over nuclear export licensing

FACED with growing difficulties in implementing its non-proliferation policies, the Carter administration is seeking to simplify its task by removing an independent check on the adequacy of safeguards covering nuclear export currently provided by the Nuclear Regulatory Commission.

The NRC itself is split over whether such a move is desirable, but it is being strongly resisted in Congress, for in passing the Nuclear Non Proliferation Act of 1978, legislators specifically invoked the NRC to try to prevent the non-proliferation debate from becoming too politicised by remaining within the sphere of the State Department and the White House.

Under the terms of the Act, the NRC is required to evaluate whether facilities to which nuclear fuel or technology is being exported contain sufficient technical and institutional safeguards to ensure that neither could be diverted to military purposes. However its decision to reject a license application can be overturned by the President — whose decision can in turn be reversed by Congress.

In recent weeks, the White House has been faced with a number of sensitive decisions on nuclear exports. Some in particular have been the direct result of the two-year time limit imposed when the Nuclear Non-Proliferation Act was signed in 1978, requiring nuclear shipments to be cut off after 10 March 1980 to all countries which had not renegotiated nuclear agreements along the lines of the Act.

Attempts to enforce controls on the export of enriched uranium to Europe,

however, were bitterly resisted by leaders of the European Economic Community in the summer of 1978, particularly because of the restrictions placed on the transfer of fuel originating in the US.

With the US administration's failure to generate a consensus behind its general strategy at the International Nuclear Fuel Cycle Evaluation (INFCE) meeting in Vienna last month, it is reluctant to provide further embarrassment.

So, last week President Carter agreed that, rather than bring an embargo into effect, western nations would be given permission to continue buying American enriched uranium as reactor fuels, even though they have not complied with the terms of the 1978 Act. The dispensation is allowed for one year, under the President's power to allow exemptions while negotiations are taking place.

A more difficult case is raised by India, which has persistently refused to sign the Nuclear Non-Proliferation Treaty or to accept international inspection of its nuclear facilities — two requirements for being granted export licenses under the 1978 Act — as being unwarranted interference in internal affairs.

The Administration is being pushed by Congress and environmental groups to maintain a firm line in refusing to grant export licenses covering a shipment of 21 tons of uranium for nuclear power plants at Tarapur in India. Under the terms of the Act, nuclear shipments are now banned unless the President determines otherwise.

However, President Carter is thought to

be actively seeking ways of not offending Mrs Gandhi, indicating last week that he is now merely seeking assurances that the fuel will not be used for explosives, rather than any broader commitment — a decision that would probably be accepted by Congress.

In a separate domain, the US is actively engaged in trying to persuade Switzerland not to export to Argentina a facility capable of producing heavy water, which the administration feels could, as in India several years ago, be used to help produce weapons grade plutonium.

Officials in Washington are said to be worried not only about the fact that this would make it easier for Argentina to construct its own nuclear weapons, but also about what they see as a lax attitude by the Swiss authorities in demanding looser safeguards on nuclear exports than had been agreed in 1978 among the main supplier countries — including Switzerland.

Finally the US has so far been unable to get very far in developing an aid package acceptable to the Pakistani government which would both guarantee Pakistan adequate protection against any Soviet threat but also persuade it to give up its attempts — still denied by Pakistani officials — to produce a nuclear bomb.

Faced with these interlocking problems, Mr Gerard Smith, Assistant Secretary of State for non-proliferation issues, has been actively lobbying Congress for a change in the legislation that would remove the NRC from, as one state department official puts it, "second guessing the President on foreign policy issues". □

The Netherlands

In an energy memorandum designed to foresee the country's needs until the beginning of the next century, the Dutch government has recommended a drastic reversal of its post-war policy to run down coal production. From a pre-eminent position in 1955, when coal contributed 96% of power plant fuel, the mining industry declined as natural gas fields came on stream in the sixties and early seventies. By 1975, coal contributed only 0.7% of the country's power plant fuel stock, and even after the 1973 oil crisis (which led to the re-opening of mines in the south of Holland), it still contributes no more than 6%.

Now the government intends to boost coal production to a level where it supplies 40% of power plant fuel (12 to 14 million tonnes of coal a year) by the year 2000. Gas and oil, which currently account for 90% of power plant fuel, will be restricted to a 20% contribution, and a decision on where the remaining 40% can be found will be delayed at least until 1983.

At present there is an over-capacity in Dutch electricity generation, but this will disappear gradually as older plants are taken out of operation. Of the country's total of 15,000 MW installed capacity, 520 MW is nuclear, and although the government made a decision in principle six years ago to install three more nuclear

Dutch to fall back on coal

plants, discussions on specific proposals are expected to occupy at least two or three years yet.

An increased role for coal in industry is also foreseen in the memorandum, which talks of fluidised bed combustion accounting for the use of five million tonnes of coal in the year 2000. This is expected to require an investment of at least 1½ billion guilders (£1 = 4.4 guilders).

Eventually, new technologies of *in situ* gasification and hydraulic mining will be needed to exploit the 100 billion tonnes of coal lying beneath the entire country at currently non-viable depths of 1,000 to 6,000 metres. In the meantime, an investment of three billion guilders will be required for the gasification of 7½ million tonnes of coal annually for industrial fuel and chemical production.

This coal gas will also be used in a mix with high-calorific North Sea or Algerian gas, or even burnt in power stations — an area in which Shell and the utilities have already taken an initiative.

In a separate report, a group of Dutch industries has concluded that the role of

coal should be even bigger than the government intends, and they call for an expenditure of 25 billion guilders (in contrast with the official estimate of 10 billion) to get the coal industry into gear by the turn of the century. They want reserves to be re-exploited at a quicker pace, and they urge demonstration projects in gasification and fluidised bed combustion within the next two years. To see this combustion revolution on its way, the industrialists suggest, Holland should have a national energy fund and a special under minister for energy (which at present remains in the portfolio of the Minister for Economic Affairs).

A corollary of both the government's and the industrialists' dreams for Dutch coal is that research and development will be required to ensure that the environment remains as clean as it is currently on a gas-dominated energy regime. Anticipating an increase in coal use from 1½ million tonnes now to 26 million tonnes a year by the end of the century, the government is setting up a national R&D programme to concentrate on atmospheric problems, pressurised fluidised-bed combustion and coal gasification, while applications will be explored for flue gas desulphurisation work already carried out elsewhere.

Altogether, the research effort will cost 750 million guilders over the next five years.

Caspar Schuurings

International exchange

Scientific forum was not a great success

LAST month's Hamburg Scientific Forum was probably the first and last of its kind. For while the two-week meeting was useful as a platform for western participants to air their views on human rights, and for the eastern bloc served as a reinforcement of the notion that detente is still a going concern, it was in scientific terms virtually useless.

Certainly the UK government was unimpressed by the event. Speaking in the House of Lords, Lord Trefgarne has said that while "a number of useful conclusions" were reached at the forum, he could not give any general assurance about support for future meetings.

Earlier, the UK delegation said in its closing statement that the "practices and processes of the Conference on Security and Cooperation in Europe" were hardly suitable for establishing a framework of scientific meetings. Soviet participants were so embarrassed by the human rights debate which dominated the forum that

they are likely to demand strict limits on the agenda of any future meeting. In the event they will probably be saved the trouble. Member governments meeting in Madrid later this year are likely to shelve proposals for another forum.

While it lasted, the forum brought together well-staffed national delegations of experts prepared to put their heads together in order to discuss "interrelated problems of common interest" in medicine, the human environment and urbanisation. What they produced was not always original thought, but it was interesting as an indicator of the divergent paths seen by different countries as solutions to broadly similar problems.

The energy group, for example, found that the participant countries interpreted "alternative energy" in widely different ways. To the Soviet Union it meant primarily nuclear energy, with the emphasis on fusion research. Sweden stressed energy from biosystems, while Professor P. Boger (West Germany), urged the biological production of hydrogen as a means of using solar energy.

The food production group was hardly revolutionary in its findings, and called simply for "sustained R&D efforts in all aspects of the food system". International cooperation, it decided, was particularly needed in the development of plants with higher photosynthetic capacity, more efficient capability to use available mineral

nutrients and better ability to withstand environmental stresses. In particular, international efforts are needed in identifying and preserving plant and animal germ plasm in the natural ecosystems. More comprehensive gene banks should be set up "to preserve genetic materials for the benefit of plant and animal production in the future".

The medical group, dealing with cancer cardiovascular and virus diseases stressed the need for data sharing and the avoidance of unnecessary duplication. Further international cooperation, they found, is needed in such fields as standardization of diagnostic materials, recombinant DNA (including safety regulations and the evaluation of benefits) and large-scale studies (e.g. drug trials) where insufficient patients would be available in a single country.

Large-scale surveys of this kind, with comparison of the patterns of change in different countries were also urged by the social science (urbanisation) group. They suggested a regular programme or scientific review conferences and seminars, with special emphasis on the interdisciplinary approach. Yet, not even this group, keen as it was for meetings, suggested that a further 'scientific forum' was the appropriate setting. Rather, it suggested, the proposed meetings should take place under the auspices of UNEP, the ECE or UNESCO. □

United Kingdom

US carcinogen regulations urged in UK

ANIMAL tests should be relied on to indicate whether a chemical will be carcinogenic in humans, says a trade union document on occupational cancers published last week.

"When it comes to the interpretation of test results" says the document "the UK guidelines fall far short of their US counterparts. It is stated in the US guidelines that a positive result in animal tests is to be accepted as evidence of human carcinogenicity. Not so in the UK guidelines, where it is stated that 'extrapolation is a difficult procedure'".

Produced by Britain's half-million strong Association of Scientific Technical and Managerial Staffs, the report was launched at a press conference by ASTMS General Secretary, Clive Jenkins. "Too many of our members are dying" he said. "Tens of thousands of new chemicals have been introduced to the workplace since the war. Companies have acted irresponsibly. We want exposure to carcinogens reduced to zero. We want a new approach in law."

Sheila McKechnie, ASTMS Health and Safety Officer, said that her union "was not attempting to ban all chemicals". Of 7,000 or so chemicals that have been tested so far, says the report, "only 600-800 have shown any evidence of carcinogenicity".

Epidemiological studies — or 'body counts' as the report calls them — come too late: "keeping track of the effects of new (and existing) chemicals by observation of the effects on workers — if relied on exclusively — turns the work place into a giant laboratory for the conduct of an experiment in carcinogenesis. No trade union can accept this."

The evidence is that animal carcinogens are human carcinogens, says the report. "Even if subsequent research unearths some exceptions to this rule, it must be taken as a conservative regulatory principle, that is a principle erring on the side of safety. To refuse to accept the evidence of animal bioassays means, in practice, to insist on counting bodies as the only valid method of identifying carcinogens. This is the position usually adopted by industry — especially an industry threatened with the banning of some profitable commodity."

ASTMS is demanding a comprehensive licensing scheme for toxic substances. Under the scheme:

- Only potentially useful substances should be submitted for test.
- The tests should be carried out in public or independent laboratories and financed by a general levy on the industries supplying the substances.
- Only substances which pass these tests should be licensed for commercial use, subject to safeguards dictated by the results of the toxicity tests.



Jenkins: would love health strike

● All companies handling licensed substances should be compelled to keep comprehensive records.

The process of awarding compensation should also be completely overhauled, said Jenkins. "Some firms have internal compensation. We are urging our members not to accept it. One firm — I don't want to name it — went to a man on his death bed and persuaded him to accept the firm's compensation. He would have done much better in the courts."

ASTMS recommends that 'occupational cancer' be recognised as a broad category of industrial disease, and that cancer victims should be automatically entitled to compensation if they can show they were exposed to a known carcinogen, or were involved in an occupation that clearly shows an elevated risk of cancer. The scheme could be funded by a specially floated National Cancer Insurance scheme, drawing on both public funds and general levies on industry. "There is great sympathy in the Trades Union Council for our views" said Jenkins.

Would ASTMS take strike action over

this issue? "I'd love a strike over health" said Jenkins.

However, the US regulations urged by ASTMS are themselves under pressure, as reported in *Nature* (24 January, page 320). In the latest proposals of the US Occupational Safety and Health Administration, it is expected that the list of 'category 1' carcinogens (ones where scientific evidence of carcinogenicity is strong) will have only 150 chemicals, and only 10 of these a year will be selected for 'special attention'. US industry still feels the regulations to be too restrictive, while unions are attacking OSHA for lowering its sights.

At its press conference, ASTMS was attacked on the scientific basis of its report, which claims that 20-40% of cancers can be traced to occupational causes or synergisms. In particular two of its sources have been roundly attacked: Samuel Epstein's book, 'The Politics of Cancer', and a joint report of the US National Cancer Institute, National Institute of Environmental Health Sciences, and the National Institute for Occupational Safety and Health, titled 'Estimates of the fraction of cancer in the US related to occupational factors'.

However, said Sheila McKechnie, ASTMS did not want their recommendations to stand or fall on technicalities. If the proportion was 40% or 5% it was still too much.

So how will British industry respond? "We are trying to get unions involved in testing these chemicals, to make sure they are unbiased" said Sheila McKechnie. "Initially testing will have to be done in industry; but so far we have met no resistance from companies on the pre-testing of new chemicals. It's chemicals where companies are already deeply involved where there is a problem."

Robert Walgate

Coal produced electricity the cheapest, says US economist

AN American energy economics consultant who waged a successful environmental battle against the emissions of coal-fired power stations in the early 1970s argued last week that electricity produced from coal in the late 80s will be cheaper than nuclear electricity. If he were right, break-even would only occur in the UK if coal tripled in real price over the next eight years.

The consultant, Dr Charles Komanoff, presented his views at a joint session of the Parliamentary Liaison Group for Alternative Energy Strategies and the

environmental group Green Alliance. (Parligaes is a rapidly growing association of MPs and others designed to provide a source of information in the House of Commons alternative to that of the national energy bodies in the case of electricity, largely the United Kingdom Atomic Energy Authority and the Central Electricity Generating Board.) Later, Komanoff gave evidence before the House Select Committee on Energy.

In both fora, Komanoff based his argument on an analysis of capital cost escalation in the 162 power plants built in

the US between 1971 and 1978. Of these, 116 were coal plants, and 46 light water reactors (of which 33 were pressurised water reactors — PWRs).

Over this period the coal plants suffered a real cost increase of 68%. Almost all of this increase, said Komanoff, could be accounted for by new apparatus to control emissions, such as SO₂ scrubbers (half the increase), improved efficiency in electrostatic precipitators and the like.

The nuclear stations built over the same period became more expensive at almost twice the rate. They increased by 128% in real terms.

"If we are to project these trends" said Komanoff "we need an hypothesis for the increases". In the coal case, the reason was environmental controls. In the nuclear case, "it is impossible to attribute specific costs to specific controls, as regulations ripple through the whole plant". The provision of emergency core cooling, a containment vessel, protection against fire, the provision of monitoring instrumentation, protection against seismic events — all had their effect on costs, he said.

A further hypothesis, he suggested, was that regulations become increasingly stringent to hold down total environmental costs. If the number of coal stations were doubled, there would be demands to halve their emissions. The same applies to nuclear power, Komanoff believes "to keep total risk constant, the probability per reactor of an accident must fall".

In both the coal and nuclear case, he calculated, station costs increased by 50% for each doubling of the corresponding sector.

Moreover, the lessons of Three Mile Island have yet to affect power station costs. "The Kemeny findings are not the end of the story. The Nuclear Regulatory Commission, and the Advisory Committee on Reactor Safeguards are constructing proposals which will lead to sweeping, thorough, and comprehensive improvements in reactor design." These alone would continue the upward trend in nuclear costs "and if a Welshman sees certain safeguards applied to a reactor in Virginia, he will want them in Wales too".

In the UK, coal station costs would also rise, thought Komanoff, because "at the moment you export all your SO₂ to Scandinavia and Germany" so there would be pressure for controls, but this would not be so strong as the purely internal pressure in the US — which has now had its major effects.

The net result, he said, was that a 1150 MW nuclear plant built in the US in 1988 would produce electricity at 4.74 cents/kWh. A 300 MW coal station built at the same date would produce electricity at 3.83 cents/kWh, allowing for interest on capital costs. The capital costs would be: nuclear, \$1684/kW; coal, \$945/kW.

But, said David Widdicombe, QC, who was chairing the Parliagaes/Green Alliance meeting, the crucial question was

whether there were reasons to expect the British experience to be different.

There were, said John Jukes of the CEBG. First, progress towards the coal production targets of the National Coal Board were slow; and coal will be needed for other purposes (such as substitute natural gas), pushing its price up. "The US has plenty of cheap coal, so relativities are bound to be different". But, said Komanoff, even a tripling of coal costs would only equalise coal and nuclear electricity prices.

Further, said Jukes, the UK has an appalling record on large construction sites and "until we get that right no-one will know the costs."

"Also the regulations in the US have been piled on one after the other in a not very efficient way." Britain should avoid a detailed licensing procedure like that in the US, said Jukes, where every component has to be certified.

"We won't know the costs until we place the orders for the next advanced gas-cooled and pressurised water reactors" said Jukes, backed up in this by a colleague Mr F P Jenkin, head of the system planning branch of the CEBG. "The CEBG has no commitment to build plant which are too expensive" said Jenkin.

However, a broad call was made at the meeting for the CEBG to expand on a term in its own costings, prepared for the meeting by Mr Jenkin, where the whole cost advantage of nuclear stations comes in a single item: 'net system savings'. (See table). This term contains most of the important assumptions, said Komanoff; will the CEBG not spell them out?

"There is a strong feeling in this meeting that more information is needed" said Widdicombe. "I will certainly accept that" said Jukes.

On the other hand David Pearce, of the Department of Political Economy,

University of Aberdeen, and a prominent critic of the Windscale inquiry into the expansion of nuclear fuel reprocessing facilities in Britain, attacked Komanoff's own assumptions. "The average load factor for a coal plant in the US is 55%, not the 72% he assumes. And I found his results are very sensitive to coals costs. If coal rose to three times uranium costs, his coal/nuclear balance would be reversed."

At the Select Committee on Energy, Komanoff was also subjected to close criticism. "If nuclear reactors are uneconomic" asked Arthur Palmer "why do the utilities continue with them?" Because, said Komanoff, they feel they are fighting a holy war. "But does that eliminate the profit motive?" asked Palmer. There is no profit motive, said Komanoff, because utilities' profits are guaranteed by state law, which guarantees an electricity rate which will cover running costs and give a return on capital investment.

Komanoff recommended that if Britain were to have an expanded nuclear power programme, it should not chose the PWR. There are more PWRs in the world than any other reactor type; "and if you leave yourself open to design reviews of PWRs the world over, you will constantly face escalating costs. On the other hand if you go for a minority reactor . . ."

"If I wanted to destroy the CEBG's nuclear programme" said Komanoff "I would recommend a large Westinghouse PWR". The average load factor of the 13 such reactors over 800 MW in the US was 44% last year. "And over 50 reactor-years the average load factor has been only 52%."

Smaller coal plants, of 200-400 MW, said Komanoff, should be able to operate at 70-75% "if the electricity supply system were properly planned".

Robert Walgate

Table: net effective costs of new stations (March 1980 prices)

	Nuclear			Coal Fired		
	(a)	(b)	(c)	(a)	(b)	(c)
Capital cost of station including initial fuel (£/kW) in the case of nuclear	850	1000	1150	416	490	564
Net effective cost (£/kWpa)						
Capital charges	71	84	97	35	41	47
Decommissioning	2	2	2	0	0	0
Other operating costs	12	12	12	10	10	10
Net system savings	-121	-121	-121	-30	-30	-30
Total	-36	-23	-10	+15	+21	+27

Notes:

- For both nuclear and coal fired stations, (a)×(c) represent -15% and +15% respectively on the central estimate of capital cost.
- Capital charges include those on associated transmission works.
- Capital charges include the annuitisation of interest during construction.
- Decommissioning costs include plant scrap value allowance.
- Other operating costs cover salaries and associated costs, repair and maintenance, rent, rates, insurance, etc.
- Net system savings are system savings after deducting station fuel cost.

NEWS IN BRIEF

Aid recommended for nuclear test victims

NUCLEAR weapons testing probably resulted in "a small number of cases of death or disease" among those living downwind of the test sites in Nevada during the 1950s and early 1960s, for which the government should be prepared to take responsibility, according to press reports of an investigation carried out by a White House study group.

A local paper in Nevada, the *Desert News*, said that the study group's 57-page report, currently under "active consideration" according to White House officials, recommends that legislation should be drafted to compensate those who may have developed cancer as a result of the testing.

However the group is said to oppose legislation proposed by Senators Edward Kennedy and Orrin Hatch conceding liability for radiogenic cancers among residents of the fall-out areas during the period of the tests with courts directed to set the damages payable. The group argues that a more restrictive proposal should be worked out.

Over 950 current and former residents of states surrounding the testing region have already filed over \$2 billion in claims against the government for injuries which, they claim, are directly related to the testing.

Inflexibility of Delany Clause criticised

DR DONALD Kennedy, until recently commissioner of the Food and Drug Administration and currently provost of the University of Stanford, last week criticised the Delaney amendment — which forbids the use of any chemicals in food shown to have caused cancer in laboratory animals — as being too inflexible and in need of revision.

Dr Kennedy said that the clause in the FDA's authorising legislation codifies the hypothesis that there is no threshold

concentration below which a chemical does not cause cancer. And although this hypothesis "probably holds most of the time", Dr Kennedy said that he was "as certain as I can be of any scientific prediction that some day, very soon, some compound will be demonstrated to have a threshold level for cancer causation."

The Delaney clause was a good version of how society should deal with such risks, he said, but was suffering because of its inflexibility, leaving no room even for a convincing scientific demonstration that there is a safe level for some cancer-causing substance.

"Any law that purports to deal with science ought to leave room for scientific progress," said Dr Kennedy. "For that reason I favour altering the Delaney clause and similar provisions to reflect presumption rather than certainty — a presumption of risk that could be rebutted by a scientific showing that a certain level is indeed safe, or that the use of a particular kind of experiment actually overstated the risks."

Cracks discovered in reactor turbine blades

CRACKS have been discovered in the turbine blades of ten nuclear reactors constructed by the Westinghouse Corporation. Company officials point out that, since the turbines operate totally outside the nuclear plant, they are not considered to present a major source of safety concern at present.

Critics, however, claim that if a turbine disintegrated as a result of the cracking, fragments of steel blades could seriously damage the reactor vessel or a safety system. The Union of Concerned Scientists has asked that plants where serious cracks have been found should be shut down immediately for inspection.

The US Nuclear Regulatory Commission has identified cracks in the turbines of ten reactors, and has drawn up a list of 19 reactors that should be inspected for cracking during the next scheduled shutdowns. The commission is currently studying whether an earlier shutdown is warranted for any of the reactors.

ESA countries approve Spacelab cost overruns

THE member states of the European Space Agency voted unanimously last week to reaffirm their commitment to complete the Spacelab programme in spite of cost overruns. Under the initial Spacelab agreement, member states retained the right to withdraw from the programme if cost overruns exceeded 20%. The Spacelab programme board last week approved a further funding up to a 40% over-run after which another vote can be

taken. Only Italy decided to reduce its contribution (from 18% to 1%) and the remaining ten states have assumed the increased cost. The 1980 Spacelab budget now amounts to \$304 million.

In the meantime UK space scientists have raised objections to the Science Research Council's rejection of Spacelab experiments in the last round of budget applications. "They are reluctant to recognise that space experiments are costly. If this keeps up the UK will lose its expertise in building space hardware," said one of them. An SRC official said that the SRC judged every experiment on a cost effective basis — "it's a question of value for money".

Violence, barricades and arrests at Plogoff

MILITANT demonstrators barricaded all roads into Plogoff recently — the Brittany town in cap Sizun where a complex of four nuclear reactors is to be sited — after Mme Amelie Kerloc'h, deputy mayor of Plogoff had asked the townspeople "to make the commune an island inaccessible to the police". The barricades followed a week of confrontation with a 600-strong police force in which police bombarded demonstrators with teargas grenades thrown from helicopters and the demonstrators countered with Molotov cocktails and stones against massed police formations. An explosive charge damaged the Loc'h bridge and 14 demonstrators, all young workers from the vicinity, were arrested. A delegation of mayors and elected officials from the entire cap Sizun region has demanded the withdrawal of the police as a first step to defuse the situation.

US publishes science magazine in China

THE first issue of a new Chinese language science and technology monthly, *Science and Technology Review*, was launched last month in Beijing. Published by the Education and Science Society in the US, the magazine is edited by US scientists and overseas Chinese scholars including C.N. Yang, Professor of Physics at Stony Brook and winner of the 1956 Nobel Prize in physics. 100,000 copies of the first issue are being sold in 29 municipalities, provinces and autonomous regions. The 104-page first issue includes articles on comparative economic systems — Japan, Yugoslavia and the USSR — and an interview with Sheldon Weinberg and Steven Glashow winners of this year's Nobel prize in physics. According to Xinhua, the Chinese news agency, the magazine, which is the first foreign publication to be given wide circulation in China, is to be given special attention by "the decision-makers and administrators of the state".



FEATURES



Mahler's revolutionary study

THE 1970s saw a debate on the role of science and technology in the economic and social development of the Third World. The failure of the development strategies propounded during the fifties and sixties had become clear. It has been repeatedly pointed out that the benefits of development programmes are not filtering through to the poor majority; western technology often helps to perpetuate rather than reduce economic and social injustice.

No international organisation has taken this to heart more than the World Health Organisation (WHO). This is ironical, because WHO is probably the best and most respected UN organisation in terms of technical competence, and this competence lies in western-style health technology.

But WHO has now set the goal of "health for all by the year 2000", and to achieve this, it is no longer advocating big hospitals, sophisticated medical technologies, and specialist medical professions. It is now talking about

The UN General Assembly recently approved the World Health Organisation's goal of health for all by the year 2000. Here, the architect of WHO's new policy, Director General Halfdan Mahler explains his thinking to **Anil Agarwal**

concepts like primary health care which are based on such hitherto taboo ideas as community participation in health services, self-coping and self-care, appropriate technology for health, appropriate health cadres like 'barefoot doctors' and traditional medicine. The UN General Assembly endorsed WHO's new goal and its strategy of primary health care in November. All governments are now being asked to formulate by June strategies to achieve that goal.

The radical change in WHO's approach

to health care has been described as the "Mahler revolution". Director General Halfdan Mahler has been a major force in the recognition that past health strategy has failed to bring any organised form of health care to more than half of the world's population living in rural areas and urban slums.

But the attempt at change has produced resistance. There is very little real political commitment, either in the developed world or amongst the elite of the developing world, to do much for the poor. WHO now finds itself saddled with idealistic international resolutions that few countries are really keen to implement back home, whether it is in terms of coming forward with the funds required to bring health care for all, or in terms of implementing those strategies and re-ordering domestic health priorities.

Lack of a real constituency is not the only problem. Even to prepare the intellectual framework for the change has meant antagonising such powerful actors

in the existing health drama as the drug industry, the baby milk powder industry, the tobacco industry, and not least, the medical establishment. The latter exists within WHO itself. Convincing WHO's own secretariat of the new mission has not been an easy task; Mahler ruefully refers to his "electric chair".

Mahler was frank about how WHO is coping with the stresses and strains of this change.

"The constitution of WHO speaks of health as a human right; that physical, social and mental well-being should be at the highest possible level, so it was in the constitution of this organisation that it should have not only a narrow technical mission but also a social mission. This social mission was, however, neglected. In the 1950s, we had the post-war hangover, the white man's guilt complex, that something ought to be done about the poor world. Isolated technical assistance programmes — lollipops, as I call them — were developed. Everybody tried to do his missionary activity in the Third World, including me. WHO did some excellent technical work but in a number of rather narrow fields. Then came the reckoning in the 1960s that it is all very well to do a little bit of malaria control, some little work on some other communicable disease, but it doesn't get health down to the grass-roots. So in the 1960s we had agonising questions such as: can WHO assume its real social mission?"

Mahler's strategy to bring the social mission to the fore has been simple: "I felt that the most important thing would be to invite the member-states to take their responsibility, whether they came from the North or the South. The big words, the collective decisions in the political organs of WHO, could no longer remain dead letters. This meant that the member-states had to ask WHO to be their bad conscience. We challenged them with the idea of health for all by the year 2000. It was not shot down in flames. The view was expressed that this indeed ought to be our aspirational target. This then raised the question: what should be the tool in moving towards that goal? That led to the Alma Ata conference on primary health care (PHC) in 1978 and the charting of the primary health care approach."

"There is now", according to Mahler, "in the governing bodies of WHO a realisation of what this movement towards health for all could become. Member states, however, often tend to forget individually what they have decided collectively".

The health debate has focused on two types of contradiction, says Mahler: "First, the conceptual contradiction. Previously we never dared to say that the end-product, health, should be attainable by everybody. We only spoke of services. We said we will try to get as big a coverage of services as possible.

"We have only a marginal degree of

popular participation in health care. This is the fundamental contradiction: if health doesn't start with the individual, the home, the family, the working place, and the schools, then we will never get to the goal of health for all. Even if we take the example of the industrialised countries, self-care, self-responsibility, self-coping in the individual, family and community, represent 50-60% of all care. Unfortunately, health professionals are rarely willing to trust people to such an extent that they acquire power to make the decisions that have to do with their own health. I just don't believe that anything in the North, for instance, can truly be changed when it comes to overeating, overdrinking, abuse of drugs, overstress

‘If health doesn't start with the individual, the home, the family, the working place, and the schools, we will never get to the goal of health for all’

and alienation, if we do not give power to the people to understand their own predicament, and then make conscious decisions about where they want to go."

In every country there is this fundamental contradiction between people and the medical consumer markets that they are provided with. "Of course", says Mahler, "the consumer markets in the developed countries are better because there you find, for instance, drugs all the time. That is because that is where the profits are. But if you go to a developing country, you find drugs available only for 3-4 months a year, which does not exactly generate confidence in the health services. Now by first looking at people's needs we know that we can no longer talk about putting drugs at random into the market. We must have a national policy whereby certain drugs are available all the time. Then we know we can get penicillin for the kid whenever he has pneumonia."

The second major contradiction is that the health services like to deal with only the top 10% of the health pyramid. Medical education is such that it teaches the new professionals to love the highest technology and medical research and development is such that it likes to generate only the highest technology. Resource distribution in the medical sector is such that 80-90% of the resources go to meet 10-15% of the health problems."

In his speeches to the World Health Assembly, Mahler has repeatedly emphasised the need for a strong political commitment if these contradictions are to be resolved. They can be found everywhere in the world. "I have been in two developing countries in the last few months," explained Mahler. "In one of them, the head of state asked me for help in fighting the vector the mosquito. I replied:

'Mr President, indeed we have been trying to do that, but I do not think that we have vector control programmes that can prevent what is probably more important to you, that is, 100,000 or more kids dying from malaria every year. So by all means fight the vector, but let us not forget that it is somehow more important for people that 100,000 of their children will die this year from malaria.'

"Now this is not a play on words. By focusing on the vector, we could easily bypass our political responsibility. (The first thing that ought to be done is to organise a system through community participation, if necessary, by which everyone who has or is likely to get malaria can easily get a tablet to prevent it or cure it).

"In the other country, the prime minister asked that WHO's entire budget for two years be used for buying a body scanner. For the same money that country could immunise its children against measles thus saving half a million children dying in the next ten years."

But what is actually happening in the health ministries of the Third World? "I am not speaking of any miracles happening, but I am convinced that there is a reappraisal going on in many developing countries. This is setting into motion within these countries strong confrontations." WHO's role, as Mahler sees it, "is to enable member-states who find it politically difficult to cope with these confrontations to make use of WHO's neutral platform to cope with them.

"Look, for instance, at drugs. It was very difficult to get a national drug policy going in the developing countries because of the resistance, whether it came from the medical specialists or the general physicians. WHO, for better or worse, has been trying to get the best talent around the world and they have been saying that a concept like essential drugs is perfectly valid. Essential drugs that can cope with 95% of the problems even in relatively sophisticated societies number around 200. But for the Third World villager and urban slum-dweller great miracles can be achieved with 50 well-chosen drugs. It was WHO's neutral role as an arbiter that led to the realisation that there is nothing shameful in speaking about essential drugs. Now the onus is back on the governments. Can they now get their experts together and enforce this concept?"

In the confrontation over the marketing of baby milk powder, Mahler believes "WHO has played a unique role by providing a neutral platform. What we achieved was not a miracle, but when you compare it with the situation ten years ago, it was remarkable that industry, non-government organisations, and member states from developing and developed countries, could arrive at this kind of consensus. It was certainly a very



Halfdan Mahler:

‘Medical education is such that it teaches the new professionals to love the highest technology and medical research’

unpleasant kind of confrontation. I was very concerned about WHO getting involved in it. But slowly we picked up the courage and hoped that it would not end in a disaster. And it didn't. Now member-states will have a code of conduct for marketing of infant formulae. This will be a model which they can adapt to their conditions.”

WHO has also had confrontations with the scientific community, which Mahler says “finds it very unpleasant when you speak about social relevance and social equity. But take these one million kids dying in Africa from tropical malaria. Somehow, it must be possible to mobilise enough social commitment among those scientists involved in health research to say that this cannot be tolerated when we have such fantastic scientific tools today. Well, scientists initially said we are prepared to do our little bit as long as it is marginal. Now through our diarrhoeal diseases, tropical diseases and human reproduction research programmes, we are dragging them. We have opened our doors to the scientists. I think the response has been very good and will be increasing in the future.”

But is providing a neutral platform and creating a consensus enough in these confrontations? Is WHO happy with the speed with which the essential drugs programme is being implemented, for instance? “No, certainly not,” says Mahler. “But then we have to ask ourselves whether WHO isn't becoming a little megalomaniac in this field of drugs. WHO now finds itself fighting the world at large. We are now moving straight into technology, production, patents, trademarks, the elements of a new international economic order in the widest possible sense. Is this WHO's role? I think we have set the scene, and other organisations like the UN Conference on Trade and Development, the UN Industrial Development Organisation, World Bank, etc, should now play the key role. If you

look at the drugs scenario, despite its imperfections, there is already a lot of courage, in espousing the essential drugs idea, courage by a number of developing countries now developing national drug policies, and courage by member-states in the Pacific region, for instance, who have got together under WHO auspices to decide on collective purchasing so that they can negotiate with drug suppliers much more strongly than in the past.

“But still I am not happy with the speed of the essential drugs programme but that is because I am not happy with myself. I don't seem to have the necessary social or economic ability to accelerate the process. And, therefore, I want to involve UNIDO, UNCTAD, etc into whatever I can do, and the drug industry itself, by asking it what can be done. I can give you an example. It was rather startling. We had a conference and the industry spokesman said that the production of vaccines and sera is so competitive that they were losing interest in it. What conclusion can you draw from that? When you want health for all and want to prevent six million children from dying from tuberculosis, whooping cough, diphtheria, measles and polio every year, then you need to vaccinate 100-120 million children every year. In order to get that you need a vaccine price that is low. But when the price is low, you can't get the products. So my conclusion is that we can no longer treat these vital, essential components of people's health as normal commodities in the market place. They may have to be taken out of the market place and other ways found to produce these essential drugs. But all of this obviously generates tremendous contradictions.”

But what about the contradictions inside the house? To what extent has the Mahler revolution filtered down within WHO itself? Mahler has several times told his staff that WHO should not be an academy of health sciences but a centre for managing health.

Mahler agrees that there is an internal contradiction. “It is difficult to take the plunge from doing things that you do well, to continue to do them well but with a very much sharper relationship with this overall target of health for all. The secretariat is asking, are we not throwing out the baby with the bath water? We had all these good technical things. Is the gap not too big between all these expectations of a social mission and our limited but well-done technical mission? This challenging target of health for all set by the member-states collectively is probably the only way of bringing the high priests on their toes — the high priests inside the secretariat and in the individual member-states.”

Could not WHO's technical reputation suffer in the process? Could it not find that it has compromised its technical reputation as well as failed to achieve its social goal? Mahler believes that the danger is there, but it does not inhibit him.

“Take the case of smallpox eradication

which was first decided on in 1957. The internal technicians in WHO — and don't forget I am one of them — claimed for ten years on technical grounds that the job could not be done. However, in 1968 we turned the argument around and said smallpox shall be eradicated and the political commitment shall be taken for granted. We worked backwards to see how our techniques and our management would allow this political commitment to manifest itself, and how our technologies could do the job. And we succeeded.”

“Most people analysing the contemporary world will say, forget about health for all. They will tell you that you will soon have a nuclear war in several of the developing countries. If you look at the world without a feeling for this imperative, historical necessity to change, then obviously you will generate a climate of gloom, and then be discouraged by it. Inside WHO, we could have gone on with basic health services, some tuberculosis, some malaria work. But now we have forced ourselves politically into interfacing with the collective aspirations of our member-states.

“I do not think that there is one person who can say that this change requires less of him technically. I think it requires infinitely more scientific imagination, in order to lift our technological thinking out of our cocoons in such a way that it relates more meaningfully to this broader social perspective. But I do not believe that in the process any of the conventional technical activities will be lost. In fact, the large majority will have become infinitely better within the next 5-10 years and more relevant to our social targets. In recent years, we have been generating the various research programmes, the expanded programme for immunisation, and planning for clean drinking water and sanitation, with the target of ‘clean water and sanitation for all by 1990’. All these programmes have highly explicit objectives. These objectives are key components of primary health care. The programmes are helping people somehow to identify with the components of the overall strategy. So now you have within WHO a progressive improvement, from a kind of technocratic alienation to begin with, to an identification with the major programmes that will be filling out the various major gaps in primary health care.”

WHO must become more responsive to the realities of the world, he says. “I have been asking member-states: do you really think that a technocracy like WHO should continue to exist? Either you believe that you can make WHO big enough to assume this brokerage role at the global, regional and country levels, or you don't. If you do, then we have to make a lot of agonising changes, and there will be no peace. But if and when the member-states realise that WHO cannot assume that larger mission, they should have the courage to say that we

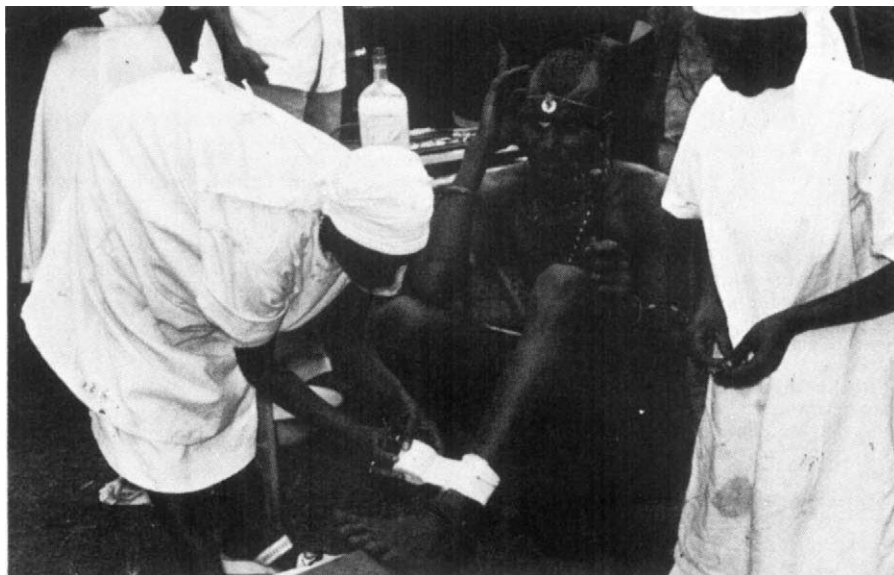
couldn't make the organisation swing around to what we all believe should be done, so let's get rid of it. Let us not keep the system for some spurious political motive."

But Mahler is optimistic on several counts: "I think that we have to a large extent successfully called the North-South bluff in health. And having called it, it is beginning to hurt. I know many politicians in the Third World are very uncomfortable with the present situation in which the elite is eating up 80-90% of the resources. There is a genuine debate on health in the industrialised countries too, though to a lesser extent."

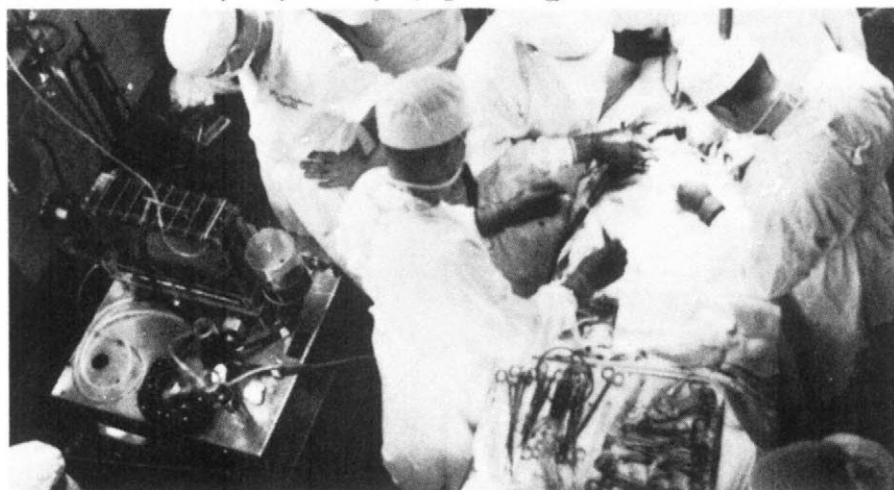
Given the "hysterical conservatism" that is rampant in the developed countries, Mahler is still not sure to what extent they are prepared to help the developing countries achieve the goal. But he argues the South "now has a much better chance of moving towards self-reliance in health. Now it can challenge the North by saying that to achieve our collective goal of health for all which we all have collectively decided, we will do 95-97% of the job ourselves. Will you not let us have that 2-3% of support that is essential for us to make the best of our 95-97%?"

And why should the developed world help? Because in the process it would only help itself. "Take the case of smallpox eradication. Ironically, it is the rich world that is making, according to a moderate estimate, a profit of about a billion dollars a year — savings in vaccine production, in treatment of complications, in surveillance mechanisms, in the tremendous waste of time involved in having to make vaccination compulsory. Now I should have thought it moderately reasonable on my part to say that half of the profits should go back to the developing countries. They did it really for the sake of the developed countries. Take Africa. Their form of smallpox was not very serious. They could live with that. If you have a million children dying from malaria, it is not the most important thing to eradicate smallpox. But because of its commitment to global solidarity, Africa had to eradicate smallpox. It just shows how inter-dependent the nations of the world are in health."

When developing countries undertake research into their own problems, even then the West benefits. "Take my own little experience in tuberculosis," says Mahler. "In the 1950s, when I went to work in India, we were exporting the classical European notion of TB sanatoria. India then had three million cases of TB and little money to diagnose, treat or immunise. Well, the government of India picked up the courage to do research itself, and indeed the consequence was that the chemotherapy research that came out of the Madras TB research centre became internationally the most important. The centre succeeded in standardising the procedure of diagnosis, and in developing the concept of



Medicine in two worlds: primary care in Africa, high technology in the North



intermittent home-based treatment. All of this research was clearly geared to India's absolutely incredible problem, and these were studies that would never have been done anywhere else in the world.

"Yet, ironically, the great beneficiaries of all this research were the rich countries. Hundreds of millions of dollars have been gained in the developed countries by making use of the Indian research. A European country I recently visited has been able to avoid constructing 3,000 beds for TB treatment because they now have the WHO expert committee's recommendations based on research done in India that home-based treatment is adequate. This was the effect in just one developed country."

But probably what the developed countries will gain most is knowledge about how to handle their own health problems by emulating the approaches being developed in the Third World — a kind of reverse transfer of technology.

"Global standards of health and well-being are declining: life expectancy, after reaching a peak, is now again decreasing; cancer rates are rising; cardiovascular diseases are rampant; drugs, alcohol, cigarettes and traffic accidents nowadays

kill more people than did all the epidemics together in earlier centuries; the aged are overwhelmed with diagnostic tools and abstruse technology, but their psycho-social and mental well-being is left largely unattended and uncared for.

"This uneasy feeling about today's medicine is widespread. As a consequence of the present high technological pitch of diagnostics and therapeutics, the very attempt to diagnose and treat one disease may produce another. As many as 20% of hospital admissions are reported to fall in this category in certain parts of the affluent world. Our world has become too accustomed to believe that with a little bit of resources and a little bit of knowledge any problem can be solved.

"I think that an appreciation of the South's real health problems will have a better chance to emerge once the North gets a feeling for the incredible irrationality that exists in its own health system. This will be an additional factor leading to the true global solidarity necessary to achieve health for all the world's peoples. If you look at our concepts about the transfer of science and technology, then I have no hesitation in saying that we really are out among the *avant-garde*."

CORRESPONDENCE

Two good reasons to reject the PWR

SIR, — Sir Alan Cottrell (*Nature*, 28 February, page 804) has put his finger on the principal weaknesses of the PWR — the two-phase coolant and the pressure vessel failure problem — either of which is a cause for doubting the wisdom of adopting this reactor for the UK. Taken together they are sufficient reason for abandoning the idea completely. One might wonder why our electricity authorities are so keen to get into the PWR business at all.

Power station construction engineers are perhaps the main pressure group. They have not performed well with the AGR for many reasons, including starting construction long before the design was ready, and trying to develop and prove major components as they went along. The resulting delays and cost increases are well known, and technical difficulties have been introduced that have caused poor operating records in the early life of finished stations. Now, with remarkable lack of introspection, the same engineers are identifying the culprit as the AGR itself; amongst other things, they say, it is too complicated mechanically and needs too much work on site. They regard PWR as the solution to their problems (or should we say shortcomings), for they have convinced themselves that it is much easier to build than the AGR. This is the dangerous sort of complacency that led to the Dungeness B fiasco, which was very nearly repeated with SGHWR.

The AGR is based on sound principles, uses good materials of construction, and has excellent safety and operating characteristics. This is a recipe for success but it does need first class detailed design and project engineering. So does any other type of reactor. It is no good running away from the gas-cooled reactors because it is thought, wrongly, that the PWR is an easier option. This simply wastes our limited technical resources to promote a reactor with some frankly undesirable properties. The better course is to recognise our failings, capitalise on experience and concentrate our best professional efforts on making AGR the successful UK thermal reactor that it can now be.

Yours faithfully,

DR. C. P. HAIGH

Painswick
Old Sneed Park
Bristol 9

Dr. C. P. Haigh was founder and for 14 years Director of the Berkeley Nuclear Laboratories of the CEBG. Subsequently he was responsible for overseeing the design of CEBG generating plant, including the nuclear reactors

Concentrating ethanol

SIR, — The amount of energy used in distilling off ethanol has not escaped attention to quite the extent that Taylor (*Nature*, 17 January, page 714) suggests, (see E.V. Anderson, *Chem. Engng News*, 31 July, 1978, page 8). In a simple still, nearly as much energy is discharged in the cooling water as is available from the ethanol. This waste of energy can be greatly diminished, at a price, by using multiple effect or vapour compression distillation units. Vapour compression is particularly attractive because ethanol could then be continuously removed from the fermentation mixture at low pressure so that its inhibiting action on yeast fermentation would be minimised.

There are various alternatives to distillation. The most interesting is the use of a semi-permeable membrane. Thomas Graham, in the paper (*Phil. Trans. Roy. Soc.* 151, 183, 1861) in which he introduced the word 'colloid',

mentions "the well-known bladder experiment of Sömmerring". He gives no reference, but explains that the permeability of bladder is selective and water evaporates from the outside so that more concentrated ethanol remains within. This may well have been a traditional technique in Europe.

Geoffrey Gorer (*Himalayan Village*, M. Joseph, 1938) says that beer is hung up in pieces of gut in Bhutan so as to make a more potent drink. Perhaps we still have something to learn from primitive technology.

Yours faithfully,

N.W. PIRIE

42 Leyton Rd, Harpenden

Soviet biotechnology

SIR, — Recently there has been a number of comments in *Nature* about biotechnology. In the issue of 10 January, (page 123) there was an attempt to assess the world situation in which, for instance, Japan was quoted as a world leader in this field; we would not dispute this claim. However, that "most socialist countries . . . are somewhat secretive as to detail" is a slightly misleading conclusion.

Whilst we were preparing the Society for General Microbiology Symposium on Microbial Technology in 1979 it came to our attention that the Soviet Union and other communist bloc countries also were becoming increasingly active in biotechnology. In 1978, for example, President Brezhnev at the 25th Party Congress cited the microbiological industry as a key growth area in the Soviet economy, the aim during the current five-year plan being to develop microbiological industry four times faster than any other sector of industrial activities.

The degree of commitment to biotechnology in the USSR can be judged by reference to just one programme, namely single cell protein (SCP) production. In 1977 Academician N.M. Zhavoronkov announced that the production of fodder yeast from various internally available raw materials was being planned to make the USSR self-sufficient in protein feedstuffs for animals by 1990. The 1980 forecast for SCP production in the Soviet Union is of the order four million tons per annum, a figure that makes interesting comparison with the UK's largest programme, the ICI "Pruteen" project, which should come on stream this year and is expected to produce 50,000 tons per annum.

Yours faithfully,

ALAN T. BULL

Department of Applied Biology, UWIST

DEREK C. ELLWOOD

Centre for Applied Microbiology and Research, Porton

COLIN RATLEDGE

Department of Biochemistry, University of Hull

Sakharov's plight

SIR, — The recent action of the Soviet authorities against our distinguished colleague academician Andrei Sakharov provides another testimony about the increasing violations of human rights and civil freedoms in the USSR.

One can understand (though not agree with) the decree of the Presidium of the Supreme Soviet of the USSR depriving Professor Sakharov of his numerous decorations. For scientists, however, it is difficult to understand the meaning of the decision of the Council of Ministers of the USSR, by which all the state prizes Andrei Sakharov was awarded in the last 25 years, have been taken away from him. Is he supposed to pay back the substantial amount of money which accompanied each of these prizes? Does the Soviet government really think that its decree will make Sakharov's discoveries (like the physical

principles for fusion reactors) non-existent or even invalid? Or that Andrei Sakharov will, by the decree of Soviet government, cease to be the author of them? For the first time Soviet reality has surpassed Orwell's visions.

Internal exile to Gorky deprives Sakharov of one of the main rights and privileges of the members of Soviet Academy of Sciences to have the conditions for carrying on his scientific research. It is not clear yet, whether the Soviet Academy is going to propose the expulsion of Sakharov from its membership, but it should be made quite clear in advance that such a step would have most harmful consequences upon international collaboration and contacts in science.

Yours faithfully,

FRANTISEK JANOUCH

Research Institute of Physics, Stockholm, Sweden

Disposal of carcinogens

SIR, — The International Agency for Research on Cancer (IARC), with the support of the Office of Research Safety of the National Cancer Institute, has recently undertaken a programme of research on the disposal of laboratory wastes containing carcinogens, which will give specific instructions for destruction and disposal of the various carcinogenic wastes from laboratories.

We give here a brief résumé of a meeting of the working group responsible for establishing guidelines and priorities for the programme. We welcome comments, information related to similar work being carried out in other laboratories, and suggestions for active collaboration in the programme. As resources available for work of this nature tend to have a lower priority than original work on carcinogenicity, it is particularly essential that available knowledge is pooled.

In discussing a general strategy for control of carcinogenic waste, the working group:

- emphasised that experiments involving carcinogens should be designed in such a way as to minimize the quantities of hazardous material used;
- recognised that the scale of operations varied with different types of experiments;
- stressed that a waste treatment should yield for disposal products having minimum adverse biological or environmental effects;
- recommended that control techniques should, as far as possible, be carried out at the experimental site to minimize transportation, which may involve a public risk; and
- insisted that an experimental plan should contain the action to be taken in the event of a substance escaping from control in the course of an experiment.

A waste control strategy must take into account both the types of material for which appropriate treatment would be required, as well as their comparative importance in terms of risk of exposure during manipulation.

In discussing possible approaches to waste destruction, it was recognised that properly designed incineration techniques require careful attention to a number of variables, such as temperature of operations, air supply and feed rate. An alternative approach is to develop suitable chemical reactions which yield non-hazardous products.

From a list of carcinogens to be considered in a destruction programme, aflatoxins, nitrosamines and polycyclic aromatic hydrocarbons were selected as the immediate priorities for investigation. Alkylating agents, halogenated compounds, aromatic amines and hydrazines were also considered as high priorities.

Yours faithfully,

E.A. WALKER, M. CASTEGNARO,

Unit of Environmental Carcinogens,
International Agency for Research on Cancer,
150 Cours Albert Thomas, 69372 Lyon Cédex 2, France

NEWS AND VIEWS

View of Venus from Moscow

from D.M. Hunten

IN this issue of *Nature* appear two papers (pages 243 and 244) reporting results from the Venera 11 and 12 entry probes, which completed their missions in December 1978. Earlier that month, four Pioneer Venus (PV) probes had also entered the atmosphere, the bus which carried them made measurements of the upper atmosphere, and the Orbiter commenced a long series of measurements expected to continue at least through 1980.

The Soviet and American probes made many measurements in common, as well as several unique to each. The current reports are in the latter category, and are of particular interest in complementing the Pioneer data as published in the 23 February and 6 July (1979) issues of *Science*.

Water vapour

Water vapour, an extremely important molecule, is difficult to measure reliably in small quantities. The PV mass spectrometer has not reported any measurements of H_2O except those stemming from the cloud drop that lodged on its inlet tube. The vapours from its decomposition and evaporation left little doubt that the drop was concentrated H_2SO_4 solution, but made it impossible to determine the corresponding abundances in the free atmosphere. The gas chromatograph reported rather large (in the context of Venus) amounts of water vapour, 0.5% at 44 km, a few kilometres below the cloud base. Previous analysis of Venus' radio emission observed from Earth, had set an upper bound of 0.1%, which pertains to the

bottom 10–20 km. The paper by Moroz *et al.* (this issue, page 243) reports far smaller amounts however, using optical spectrographic techniques.

The principal investigator, V.I. Moroz, is a well-known planetary spectroscopist, and the experiment is elegant both in its power and its simplicity. It consists of a filter-wheel spectrophotometer covering the wavelength range 430–1,170 nm with the very modest resolution of 30 nm. Included in its coverage are several bands of H_2O , which can be measured with assurance that most, if not all, of the absorption takes place in the free atmosphere. A serious complication arises from the presence of an optical 'cavity' between the clouds and the ground; some of the light seen looking upward has made one or more round trips over this whole path. With many measurements during descent, the effect can be allowed for, and the height profile of H_2O obtained. The mixing ratio is reported to be 20 p.p.m. (parts per million) near the surface, rising to 200 p.p.m. in the clouds. An earlier experiment on the same principle, aboard Venera 9 and 10, gave similar results but with considerably less assurance. It is also reported that a gas chromatograph on Venera 12 gave an upper limit of 100 p.p.m. below 42 km.

I, as a spectroscopist, am strongly inclined to believe the low abundances. Needless to say, the PV experimenters, led by Vance Oyama, are concerned about the discrepancy and have worked hard to resolve it. They can find no problem with their experiment, and point out that the amount of H_2O they observed would require a large and improbable source of contamination. An ordinary cloud drop, for example, would not be enough. So far there is no reasonable explanation, but

most workers tend to favour the optical results.

Other data from the spectrophotometer bear on the structure and properties of the clouds, and on the possible presence of a number of trace gases. Considerable absorption in the blue is apparent and could be explained by the presence of S_2 , Cl_2 , Br_2 , or NO_2 at a hundredth of a p.p.m. or less (depending on the gas). Much larger quantities of sulphur, previously advocated, are clearly ruled out.

Lightning on Venus

Before the recent encounters with Venus and Jupiter, we had no evidence about the presence of lightning anywhere but on Earth. Voyager photographed what appear to be lightning flashes on the night side of Jupiter (*Nature* 280; 794; 1979), supplementing the earlier but less direct evidence from Venus. The new paper by Ksanfomali (this issue, page 244) reports the results from the sferics detectors on Venera 11 and 12. (Sferics, a contraction of 'atmospherics', are the electromagnetic impulses from lightning discharges, commonly heard on an AM radio receiver.) Abundant signals were recorded from both probes, and a periodic modulation on Venera 11 suggests that the antenna pattern was being rotated past a distant source. On the hypothesis that the individual strikes dissipate the same energy as on Earth, the 'storm' was located at a distance of some 1,500 km. This modulation is perhaps the strongest evidence that the signals were not merely due to electrical discharge from the probes themselves.

Additional support comes from an experiment on the PV Orbiter, whose purpose was to study the properties of the plasma at high altitudes. Normally the ionosphere totally reflects all low-

Donald M. Hunten is a Professor in the Department of Planetary Sciences at the University of Arizona, Tucson.

frequency radio waves, and sferics are not observed by satellites. But there are special circumstances where energy can propagate through in the so-called 'whistler mode'. These conditions, as measured by other instruments on the Orbiter, are thought to occur fairly frequently on the night side of Venus. When they do, the plasma-wave detector sees impulsive signals of a kind not normally encountered at orbital altitude. The instrument would need much better time resolution to obtain the length or shape of the pulses, but they are consistent with an origin from lightning and no other explanation has been suggested. Taken all together, the evidence for lightning on Venus is fairly convincing.

Ksanfomaliti also suggests that lightning could generate a mean nightside brightness 10^{-4} that of the day side. This source is put forward as an explanation of the 'ashen light', a phenomenon reported by many visual observers who see the whole disk faintly illuminated when the planet is in a crescent phase. It remains to be seen whether such suggestions and reports can be reconciled with the brightnesses observed from vehicles close to Venus.

The indications are that lightning is likely to occur in any substantial planetary atmosphere. Theories of electrification are faced with the need to explain its presence under a wide variety of circumstances and atmospheric compositions. □

Hammond found that progesterone secretion from granulosa cells *in vitro* taken from small immature pig follicles could be inhibited by prolactin. Short-term oestradiol treatment also suppressed progesterone secretion. However prolonged treatment with oestrogens (for 48 h) resulted in increased progesterone secretion and reversed the action of prolactin from inhibition to synergistic stimulation of progesterone. The authors conclude therefore that "oestrogens may regulate the divergent actions of prolactin in the mammalian ovary."

The basis of the oestrogen requirement for proliferation and maturation of granulosa cells is as yet unknown. Thus the change in response to prolactin seen in the present experiments may simply reflect the maturation of granulosa cells: prolactin will stimulate progesterone secretion from granulosa cells taken from large mature pig follicles. In this respect the effects of prolactin in the pig follicle may be different from those in the human where high levels of prolactin inhibit progesterone secretion from granulosa cells regardless of the stage of follicular development.

The importance of the direct inhibitory effects of high prolactin levels on steroidogenesis by granulosa cells in follicles before ovulation is emphasised by the fact that the rare cases of ovulation in hyperprolactinaemia, in particular that associated with lactation, is associated with an inadequate corpus luteum capable of secreting only limited amounts of progesterone (McNeilly, Howie & Houston, unpublished data). This suggests that prolactin-mediated inhibition of normal granulosa cell development before ovulation may explain the inadequacy of luteal function. This is supported by the observation that drug-induced hyperprolactinaemia in the luteal phase has only a slight effect on progesterone secretion from the developed corpus luteum whereas hyperprolactinaemia induced during the follicular phase results in the formation of an inadequate corpus luteum (Delvoye *et al.* *C. r. Seanc. hebdom. Acad. Sci. (Paris)* Ser D. 279, 1463; 1974). McNatty (*Fertil. Steril.* 32, 433; 1979) has provided more direct evidence in support of this. Raised levels of prolactin in plasma and ovarian follicular fluid were associated with a reduced number of granulosa cells and a marked reduction in intrafollicular steroidogenesis, not always apparent from the levels of circulating oestrogens.

Thus it seems that prolactin may have an important regulatory role in granulosa cell development within the follicle. Although present reports suggest that oestrogens may modulate the action of prolactin, the species difference in this action has yet to be resolved. The importance of understanding these mechanisms is apparent since hyperprolactinaemic blockade of steroidogenesis may in part explain the natural contraceptive effect of lactation in women and many other mammals. □

Paradoxical prolactin

from Alan S. McNeilly

LACTATION in many mammals is associated with suppression of ovulation as a result of the failure of ovarian follicles to develop to the appropriate stage. The widely varying suckling patterns in different societies (from very frequent short periods, for example in the !Kung peoples of the Kalahari, to infrequent longer periods of suckling) are all associated with suppression of normal gonadotropin secretion and of ovarian activity, and with raised basal levels of prolactin which increase further in response to the suckling stimulus (see McNeilly *Br. Med. Bull.* 35, 151; 1979). As ovarian activity normally resumes once suckling ceases, either the suckling stimulus itself and/or the raised levels of prolactin may be responsible for the suppression of gonadotropin secretion and the consequent failure of normal follicular development.

The best evidence that the prolactin level itself may be responsible for ovulatory inhibition comes from cases of pathological hyperprolactinaemia which is the apparent cause of 20% of secondary amenorrhoeas. Unlike lactation, the raised prolactin levels in this condition may be due to an alteration in hypothalamic catecholamine turnover. The way in which prolactin itself may play a part has received scant attention; it may act on the hypothalamic-pituitary axis to suppress gonadotropin secretion, or may act directly on the ovary to suppress follicular development.

Evidence that prolactin is implicated in normal follicle development has been obtained from studies of progesterone

secretion from human granulosa cells *in vitro* (McNatty, Sawers & McNeilly *Nature* 250, 653; 1974). Levels of prolactin equivalent to those seen during the normal menstrual cycle were essential to maintain LH-stimulated progesterone secretion. In contrast, high levels of prolactin equivalent to those present in lactation and pathological hyperprolactinaemia inhibited progesterone secretion, an inhibition which could not be overcome by the addition of gonadotropins (LH or FSH). A similar situation was apparent in the mouse where follicular progesterone secretion *in vitro* was also inhibited by high levels of prolactin (McNatty, Neal & Baker *J. Reprod. Fertil.* 37, 155; 1976).

But how prolactin can have both an obligatory permissive role in progesterone secretion and an inhibitory effect remains unexplained. Studies by Veldhuis and Hammond (this issue of *Nature*, page 262) go some way to resolving this dilemma. Using short term *in vitro* studies of progesterone secretion from porcine granulosa cells they found that oestradiol 17β seems to modify the action of prolactin.

Previous studies suggest that development from a primordial to a Graafian (preovulatory) follicle is initially under the control of FSH and oestradiol. LH stimulates oestradiol and androgen secretion from the thecal layer of cells surrounding the developing follicle; FSH stimulates aromatisation of these androgens to oestrogens by the granulosa cells within the developing follicle. Both FSH and oestradiol are required for the proliferation of granulosa cells and antrum formation within the developing follicle and induce an increase in LH receptors on these granulosa cells. Prolactin is detectable within follicular fluid of all follicles and receptors have been detected on granulosa cells. Veldhuis and

Alan McNeilly is a Research Scientist at the MRC Unit of Reproductive Biology, Edinburgh.

Between the galaxies

from M.G. Edmunds

WHO could ask for a better probe of the material between the galaxies than that provided by quasars? Situated at a range of distances, some very great, the quasars give off plenty of radiation to allow the absorption spectrum of the intervening gas to be observed on Earth. The use on the largest telescopes of very sensitive light detectors capable of working with high wavelength resolution has led over the past few years to rather good data on the absorption spectra of several quasars. Particularly successful has been a detector developed at University College, London. Data on six quasars obtained with this instrument have been used to deduce the existence and properties of clouds of intergalactic material (Sargent, Young, Boksenberg & Tytler *Astrophys. J. Suppl. Ser.*, 42, 41; 1980). These gas clouds are thought to be pure remnants from the initial 'Big-Bang'.

A formidable array of statistical tests on the spectra identified two main sources of absorption. The authors concentrate their attention on the Lyman- α line of hydrogen which normally appears in the ultraviolet at 1,216 Å, but the expansion of the Universe redshifts this line into the optical region as observed from Earth provided the absorber is far enough away. The Lyman- α emission (and absorption) intrinsic to the quasar and its surroundings are of course at the highest redshift in the spectrum, but many other absorptions almost certainly attributable to hydrogen appear at redshifts smaller than that of the quasar's Lyman- α . In interpreting the origin of these lines, the observed family of lines arising from three-times ionised carbon (CIV) is important. Of the two distinct types of absorbing region which seem to be present, the first kind gives rise to these CIV lines and simply represents the outer regions of galaxies through which the light from the quasar passes on its way to us. There is good evidence that the outer parts of galaxies do contain sufficient gas at the right excitation state to provide the absorption, as the same CIV line is seen in the outer halo of our own Galaxy when looking at the ultraviolet spectra of sources nearby, but outside, our own system (for example the International Ultraviolet Explorer Satellite observations of sources in the Magellanic Clouds reported by Savage and de Boer *Astrophys. J. Lett.* 230, L77; 1979).

But the other type of absorbing region which gives rise to Lyman- α but no CIV is new, and is suggested to represent true intergalactic clouds. The physical parameters of these clouds are inevitably uncertain but the proposed sizes are 2×10^7

solar masses and a typical linear dimension of 30 kpc. This extent is about the same as a medium-sized galaxy, but the density of the cloud is very much smaller, as that of a galactic mass will be more than 10^{11} solar masses. The gas in the clouds seems to contain no (or very few) heavy elements — that is, nothing more complex than helium, as no absorption lines of heavy elements are seen at the wavelengths expected on the basis of the redshifts of the hydrogen lines. This implies that the gas is a pure remnant from the Big-Bang initial synthesis which leads to production of hydrogen and helium, but very little else. This contrasts with the gas in galaxies (and their halos) which has been contaminated with heavy elements produced by nuclear reactions in stars and their supernovae explosions. From the relative distances of the clouds as implied by their redshifts, the clouds are not clustered together in space in the way that galaxies are. This is good evidence that they represent genuine intergalactic clouds, and are not associated in any way with existing galaxies.

The discovery of a very primitive kind of object in the Universe is an interesting one, but it must be pointed out that the clouds may have comparatively little influence over the evolution of the Universe as a whole. Their number per unit volume of space at present is only about five per cubic megaparsec, fewer than the observed galaxies and very much less significant in terms of mass. Thus these new objects certainly do not supply sufficient mass to close the Universe and make it eventually re-collapse on itself — a role often proposed for intergalactic material.

Sargent *et al.*'s analysis of the thermal state of the cloud gas implies that the clouds are too hot to be bound by their self-gravitation, and must be confined by an even hotter tenuous intergalactic gas. But it is significant that such a gas could still not be hot enough to provide enough X-ray emission to explain the diffuse background of X rays observed in space. (The initial results from the Einstein observatory imply that much of this background may anyway be due to the quasars themselves). The allowed cloud parameters are apparently rather tightly constrained — in these physical conditions a rather bigger or smaller mass than is suggested gives theoretical estimates that imply they would either collapse or evaporate. It is perhaps surprising that such fragile clouds have managed to survive to the present day — presumably they represent a small fraction of clouds of a whole range of sizes which formed early on in the evolution of the Universe. The other clouds either collapsed (and perhaps combined) to give galaxies and quasars, or else dispersed to become the diffuse intergalactic medium.

Can we see any of these primitive clouds near our own Galaxy? The International Ultraviolet Explorer Satellite provides an opportunity to look at the brightest quasar 3C 273, but as yet no conclusive evidence of suitable Lyman- α absorption has been reported, although the CIV absorbing gas of our own galactic halo is clearly visible. But the number density of the clouds would predict only three clouds in the line of sight to 3C 273 — so even a complete failure to detect anything would not be too surprising on statistical grounds.

Undoubtedly this study goes a long way to demonstrating that a large fraction of absorption lines in quasar spectra may be satisfactorily explained as not being intrinsic to the quasars themselves, but due to the outer parts of galaxies and to clouds in the intergalactic medium. Just what fraction of the remaining absorption lines can be attributed to gas driven out of the quasars remains to be seen. □

Electrically conducting polymers

from Paul Calvert

FOR some time people have been looking for polymers with metallic conductivity or with superconductivity. Sulphur nitride was found in 1975 to be a superconductor at low temperatures and sparked a great deal of research which now seems to have mainly fizzled out. At the moment polyacetylene seems to have potential as a metallic conductor and is the subject of much interest.

For a material to be conducting there must be a supply of charge carriers, electrons or 'holes', and an easy, high mobility path for them to follow through the material. Thus one class of polymeric conductors consists of conjugated chains, with alternating single and double bonds, such that the pi-bonding orbitals overlap to form a path along the chain. Examples of this are sulphur nitride ($-S=N-$)_x and polyacetylene ($-CH=CH-$)_x. Alternatively, stacked planar aromatic molecules can provide an easy conduction path perpendicular to the molecular plane as in graphite. Hence the interest which developed at one time in the conductivity properties of DNA, where the stacking of the base pairs suggested that the molecule could be a 'one-dimensional' conductor.

Acetylene normally polymerises to a grey, insoluble, infusible powder with all the properties of brick dust, except price. However in 1974 T. Ito, H. Shirakawa and S. Ikeda at Tokyo Institute of Technology (*J. Polym. Sci. Chem.* 12, 11; 1974) produced thin films by exposing acetylene gas to concentrated solutions of Ziegler

M.G. Edmunds is in the Department of Astronomy and Applied Mathematics, University College, Cardiff.

Paul Calvert is a lecturer in the School of Molecular Sciences, University of Sussex.

catalysts. Films formed on the liquid surface consisting of *cis*-polyacetylene at low polymerisation temperatures and *trans*-polyacetylene at high temperatures. Electron microscopy (*Nature* **282**, 286; 1979) shows these films to be mats of fibres of about 20 nm diameter and low density, 0.4 g cm^{-3} compared with a theoretical value of 1.2 g cm^{-3} .

Shirakawa in conjunction with A.G. McDiarmid and coworkers at the University of Pennsylvania measured the electrical properties of these films (*Phys. Rev. Lett.* **39**, 1098; 1977; *Chem. Commun.* 578; 1977). The conductivity of the *trans* polymer is in the range 10^{-4} – $10^{-3} \text{ ohm}^{-1} \text{ cm}^{-1}$ which makes it a semiconductor and not particularly interesting. The *cis* polymer has a conductivity of about $10^{-9} \text{ ohm}^{-1} \text{ cm}^{-1}$. However on exposure to iodine vapour to dope the polymer with 68% iodine ($\text{CHI}_{0.22}$)_x the conductivity of the *trans* polymer increased to $40 \text{ ohm}^{-1} \text{ cm}^{-1}$. At 1% iodine the polymer took on a silvery metallic appearance and went through an insulator-to-metal transition. The *cis*-polyacetylene reached conductivities of 10^2 – $10^3 \text{ ohm}^{-1} \text{ cm}^{-1}$, 11 orders of magnitude above the pure polymer values.

A range of dopants can be used (*J. Am. Chem. Soc.* **100**, 1013; 1978) and, in the same way as in semiconductors, both p and n-type materials can be produced such that a simple diode is made. Recently experimenters have used arsenic pentafluoride rather than iodine as it is less easily lost by evaporation. If the films are stretched to three times their original length the conductivity as doped becomes anisotropic, 2,000–3,000 $\text{ohm}^{-1} \text{ cm}^{-1}$ parallel to the stretch direction and 100–200 $\text{ohm}^{-1} \text{ cm}^{-1}$ perpendicular. Because the film resembles a knitted structure it is not clear to what extent this reflects changes in inter-fibre contacts and to what extent it is due to internal structural changes. F.E. Karasz, J.C.W. Chien and coworkers at the University of Massachusetts in cooperation with McDiarmid have been looking at the effect of varying polymerisation conditions and have found increases in conductivity if films are prepared at low densities and then pressed (*J. Polym. Sci. Lett.* **17**, 779; 1979).

A wide variety of spectroscopic techniques has been applied to the doped films but there is still much to learn. R.H. Baughman and S.L. Hsu of Allied Chemical report finding I_3^- and I_2 in iodine-doped films (*J. Polym. Sci. Lett.* **17**, 185; 1979). T.C. Clark and G.B. Street of IBM say that oxidation always occurs to produce ionised chains with the counter-ions intercalated between the chains. When the charge density is sufficiently high the charges delocalise along the chain and the material becomes conducting (Preprints, A.C.S. Honolulu Meeting, 1979; *Chem. Commun.* 1066; 1978).

A number of related systems have been studied. Poly(*p*-phenylene) ($-\text{C}_6\text{H}_4-$)_x has a

Role of hydrogen in amorphous silicon

from P.G. Le Comber

SINCE 1975, when it was first demonstrated that it was possible to control the electrical properties of amorphous silicon (a-Si) by substitutional doping (see Spear & LeComber *Phil. Mag.* **33**, 935; 1976), there has been a tremendous increase in the number of laboratories throughout the world working on both the fundamental properties of this material and its possible applications in solar cells, low-cost p-n junction diodes, thin film transistors, vidicon camera tubes and in electro-photography. With the recent announcement by the Sanyo Company of Japan that it is manufacturing calculators, watches and clocks powered by a-Si solar cells, we can expect this effort to be intensified still further.

The particular form of a-Si which has aroused all this interest is prepared as a thin film about $1 \mu\text{m}$ thick by the decomposition of silane (SiH_4) gas in an r.f. glow discharge. It is now known that this material contains upwards of a few percent of hydrogen and in fact many workers (but not all) believe that these relatively large concentrations are essential for the good electronic properties of these films. In view of the potential of a-Si in a wide variety of applications it is clearly important to understand the role of hydrogen in producing and possibly limiting its desirable properties.

In a recent issue of *Physical Review Letters* (**44**, 43; 1980), D.C. Allan and J.D. Joannopoulos, at the Massachusetts Institute of Technology, have calculated the electronic states and total energies of the various configurations in which H atoms could be bonded into the amorphous silicon network. Allan and

P.G. Le Comber is at the Carnegie Laboratory of Physics, Dundee University.

Joannopoulos claim that there are limitations in previous publications on this subject and conclude that strongly interacting nearest-neighbour silicon monohydrides have an important role, particularly in the interpretation of the photoemission spectra. They also find that the energy gap increases with increasing H content, in agreement with published optical data. Their results show that the valence band recedes rapidly with increasing H content but that the conduction band remains essentially unchanged. Possibly their most interesting conclusion, however, concerns the large ϵ_y peak in the density of localised gap states that is observed in field effect and other experiments. Since it is believed that it is these ϵ_y states that at present limit the performance of a-Si solar cells, by understanding the nature of these states one might be able to reduce their numbers. Allan and Joannopoulos suggest that local fluctuations in the hydrogen content can create localised states near the valence band edge and that these could be the origin of the ϵ_y states.

It would be ironic if hydrogen, thought by many to be essentially only a positive factor in producing the good electronic properties of a-Si, turns out to be responsible through these fluctuations for limiting the performance of a-Si solar cells! This suggestion cannot be the whole story, however, since a-Si produced by thermal evaporation contains essentially no hydrogen and has an even larger number of states around ϵ_y than the films produced from silane. Nevertheless many laboratories will be interested in seeing if this suggestion by Allan and Joannopoulos provides, at the least, some clue to how to decrease the number of these states and thereby improve the efficiency of a-Si solar cells.

conductivity of $10^{-14} \text{ ohm}^{-1} \text{ cm}^{-1}$ which rises to $100 \text{ ohm}^{-1} \text{ cm}^{-1}$ with arsenic pentafluoride. This polymer is a very stable powder (*J. chem. Phys.* **71**, 1506; 1979). Poly(vinylphenylene) ($-\text{CH}=\text{CH}-\text{C}_6\text{H}_4-$)_x has a conductivity of $10^{-10} \text{ ohm}^{-1} \text{ cm}^{-1}$ rising to $3 \text{ ohm}^{-1} \text{ cm}^{-1}$ at 57% arsenic pentafluoride. This material has a degree of polymerisation of only eight, barely a polymer at all (*Polymer* **20**, 1441; 1979). Polydiacetylenes ($-\text{CR}=\text{C}=\text{C}=\text{CR}-$)_x are of interest in that the polymers can be prepared as large single crystals by solid state polymerisation of monomer crystals where the monomers are already aligned in the right way for chain formation. Despite the chains being continuous through the crystals and their metallic appearance they

are not conductive as apparently they have no charge carriers (*J. Polym. Sci. Phys.* **14**, 2037; 1976).

Simple models suggest that in long conjugated chains with all the bond lengths equal, electrons should be thermally excited into the conduction band at room temperature. This does not happen, apparently because all the bond lengths are not equal (due to the Jahn-Teller effect, for the cognoscenti) and because internal rotation about the bonds interrupts the conjugation. In addition to understanding charge mobility within the chains one needs to know how the charges transfer from one chain to another, which must partly be a function of chain length. Unfortunately this is all but unmeasurable since the

polymer is insoluble.

Polyacetylene is a tractable material because it can be produced as films; there is no prospect of further melt or solution processing. It is possible that acetylene might be induced to polymerise as a single filament in the same way as a silkworm produces silk, also infusible and insoluble. Alternatively films could be rolled up to form wires but this would hardly be cheap. More probable is that acetylene could be polymerised directly onto the surface of integrated circuits to provide controllable resistance elements. Perhaps the most intriguing aspect of this work is that it requires cooperation between theoretical chemists, physicists, polymer chemists and engineers. Communication may be difficult but the number of journals in which it can be published is immense. □

Modern refuse and ancient behaviour

from Mark Maltby

THE connection between a caribou hunting group in Alaska, Hottentot occupational debris, the contents of trash cans in Tucson, Arizona and animal bones recovered from prehistoric archaeological sites may seem tenuous but they all come under the umbrella of ethnoarchaeology, a discipline that has assumed increasing importance in archaeological studies over the past 15 years.

Ethnoarchaeology attempts to interpret the behaviour of a society by means of its material culture. By trying to understand the relationship between these in present-day societies it may be possible to reconstruct the behaviour patterns of prehistoric groups by studying their archaeological remains. There is nothing new about archaeologists turning to the ethnographic record for inspiration but most anthropological fieldwork on the other hand, has tended in the past not to record in detail the artefacts, refuse, house and settlement plans, which often provide the only source of evidence for archaeologists. Consequently many American archaeologists took it on themselves to collect data more relevant to their own studies. There is now a rapidly increasing literature on ethnoarchaeological case studies from all over the world, mostly on hunter-gatherers but also on nomadic pastoralists, sedentary farming groups and even industrial societies. Analyses of the production, form, variability, use and discard of artefacts such as stone tools and pottery have been made. Various aspects of subsistence activities, the internal

organisation of settlements, the correlation between social status and material culture and the regional patterning of settlement sites and other activity areas have been studied. Several recent studies have concentrated on food refuse, animal bones in particular, and these can be used as examples of the scope of ethnoarchaeology and its potential for helping to understand the past.

Most animal bones recovered from archaeological sites were discarded as refuse but surprisingly little is known about the processes that determined the observable patterns of bone fragments and other refuse in excavated deposits. The need for a greater awareness of such processes prompted several studies of modern rubbish disposal. Perhaps the most ambitious of these — and in many respects the most unusual — has been the Garbage Project of the University of Arizona (Rathje in *Explorations in Ethnoarchaeology* (ed. Gould) 46, University of Mexico Press, Albuquerque, 1978). In this project rubbish from several hundred households in Tucson, southern Arizona was systematically sorted. The garbage has been put into more than 150 categories representing food, drink, drugs and so on and further relevant information about each item has been noted. For example, the weight, type, cost and brand of discarded food or its containers are all recorded. The analysis of this data is incomplete but the potential of such a survey for elucidating the social patterns underlying refuse disposal in a contemporary urban community is of value to archaeologists as well as social scientists.

Other ethnoarchaeological work has examined in detail the processes of fragmentation and preservation of animal bone. One of the earliest examples was that of Brain (*Scientific Papers of the Namib Desert Research Station* 39, 13; 1969), who examined discarded goat bones from recently abandoned Hottentot villages and was able to show that certain bone fragments had survived better than others and that preservation was linked to the density of the bone and the age of the animal to which it belonged. For example, sturdy fragments such as the distal portion of the humerus survived better than the more fragile proximal portion of the same bone. He used these data to refute Dart's claim (*Transvaal Mus. Mem.* 10; 1957) that the animal bones associated with australopithecine remains at Makapansgat, South Africa, had been manufactured as tools. Brain showed that the uneven representation of these bones was more probably due to differential preservation of their skeletal parts and not to any 'hominid' behaviour. Recently a more elaborate examination of the Makapansgat data has supported Brain's findings (Binford & Bertram in *For Theory Building in Archaeology* (ed. Binford) 77, Academic Press, New York, 1977). The problems of differential preservation have often been

ignored by archaeozoologists and it is only through these and similar studies that advances have been made in our knowledge of the preservation of archaeological bone samples.

The most detailed ethnoarchaeological case study on animal bones published has been carried out on the Nunamiut Eskimo in north Alaska (Binford, *Nunamiut Ethnoarchaeology*, Academic Press, New York, 1978). The Nunamiut rely principally on hunting caribou and Dall sheep for their food. Caribou are particularly important, providing the Nunamiut not only with most of their meat but also skins, marrow and bone grease. Most of the caribou are hunted during their spring and autumn migrations through the area inhabited by the Nunamiut. After a successful kill, the complete carcass may be transported to the village or skinned and more or less dismembered at the kill site and cached in the snow to be collected and possibly butchered further at a later date. Sometimes parts of the carcass may be abandoned at the kill site and never brought to the village. Some meat is required for immediate consumption in the village but most is stored either by freezing in cellars dug into the ice or by drying on racks. Further butchery may take place then and later during cooking. Certain bones are also smashed up to manufacture bone grease; others are broken to obtain marrow and some are thrown to the dogs.

All this transforms the skeleton of each caribou into a large number of distinct bone fragments scattered over several sites. The variability in the treatment of different carcasses is also reflected in the types of bone fragments deposited on the kill sites, hunters' camps and resident settlements. By combining his observations on Nunamiut behaviour with calculations of the value of the different bones for meat, marrow and grease and their general utility, Binford was able to construct models that could predict the types of bone fragments deposited at the different sites and examine the variability caused by the complex interactions of logistic, storage, mobility, seasonal and social strategies of the Nunamiut.

How can this and other ethnoarchaeological studies be applied to archaeological problems? If nothing else, they have demonstrated the potential complexity of animal bone assemblages — a complexity that has not been appreciated by most archaeozoologists and many statements that have been made about the proportion of different species represented, the relative abundance of their different skeletal elements, the kill patterns of the animals exploited and other aspects of archaeological data must now be viewed with suspicion. Ethnoarchaeological studies are, however, of most value if their information can be used constructively and Binford, for example, has claimed that some of the hypotheses generated and tested on the Nunamiut

bone assemblages can be applied to most archaeological samples. It is now necessary for archaeozoologists to adapt such hypotheses to their data. If such

adaptations are successful, the depth and range of information obtained from archaeological animal bones will have been greatly increased. □

Compartments, 'switches', 'programs' and vertebrate development

from Jonathan Cooke

It is now well established that in insect development small groups of founder cells are set aside in a hierarchically arranged series of geographical partitioning events such that each group is subsequently solely responsible for the construction of a defined piece of the epidermal body pattern. These pattern pieces, each the product of such a 'polyclonal compartment' set up in the embryo, need not correspond to morphological units as recognised by an anatomist. They are however identical and similarly positioned with respect to each other in each individual. This finding has formed an exciting link between the study of positional information (Wolpert *J. theoret. Biol.* **25**, 1; 1969) that is, the control of cells' activities according to their relative positions within the embryo during morphogenesis, and the study of possible mechanisms for progressive restriction of cells' potentialities during early development (Crick & Lawrence *Science* **189**, 340; 1975; Morata & Lawrence *Nature* **265**, 211; 1977).

There has naturally been speculation that what has been revealed in insects, with their intricate and distinctive body surface and their suitability for the somatic genetic marking of cell clones during development, is in fact general to many types of embryo. Methods now exist for marking clones over long developmental periods by injecting

the founder cells with 'heritable' substances (Weisblat *et al. Science* **200**, 295; 1978; Jacobson & Hirose *Science* **202**, 637; 1978), and first attempts to probe for compartmentation in vertebrates have been directed at brain development. Here, a complex and exact anatomy is finally produced and there is little reason to suppose that populations of cell bodies intermingle on a large enough scale, at late stages, to obscure any earlier geographical organisation. The idea of a sequential, hierarchical series of events channelling the fates of cells early in development is a comprehensible and attractive one. What are the chances of being able to detect compartmentation in the vertebrate brain rudiment if this exists — or, more fundamentally, what meaning could the concept of polyclonal compartments have there?

Regulative embryos such as those of vertebrates are those in which the body pattern progressively becomes determined within a numerically expanding, sheet-like population of cells, and not those where lineal descent seems to define cells' fates from the outset as in the nematode worm or the leech. In one trivial sense all regulative embryos ultimately exhibit a compartmentation of the founder cells for each element of their bodies during undisturbed development. But vertebrate embryogenesis contrasts rather sharply with that of insects in the timing of the restriction process relative to the onset of overt

differentiation.

In the insect work, the dramatic outcome of clonal analysis (the exploration of development by marking all the epidermal cells coming from single founder cells picked out at particular times) has been the revelation that precise restrictions occur long before overt differences become apparent among the cellular descendants. The restrictions do not apply to which histological structures those descendants can differentiate into, but to which region of the body's pattern they can participate in constructing. Very many cell generations then intervene, during which the commitment is only expressed in subtle ways such as growth rates, or the lack of intermingling that actually preserves the integrity of the compartments. It is precisely the failure of the undifferentiated cells to invade alien pattern regions that lends the compartment concept its interest and meaning.

In vertebrate embryos however, especially in the primordial nervous system, there is a wealth of evidence that commitment to pattern construction on the part of small cell groups occurs relatively late in the cell lineage, when cell number is high, and predates hardly at all the onset of morphogenesis and the first stages of histological differentiation (for example, C-O. Jacobson *Zool. Bidr. Uppsala*, **36**, 73, 1964; Chung & Cooke *Proc. Roy. Soc. B.* **201**, 335; 1978). Certain very early events, such as the separation of the embryo from its supporting structures in mammals, are exceptions to this. It is true that the decision by a population of cells within the amphibian ectoderm to form nervous system rather than epidermis occurs somewhat earlier than does regionalisation for brain structures within that committed population. But even the former decision is certainly delayed until the embryo contains more than 10,000 cells, and the ectodermal cells concerned seem on average to be in their 15th division cycle since fertilisation (Suzuki & Ikeda,

Jonathan Cooke is at the National Institute for Medical Research, Mill Hill, London.



100 years ago

Who are the Irish? By James Bonwick, F.R.G.S. (London: David Bogue, 1880). This little work is issued as the first of a series on "Our Nationalities," to be followed by three others on the Scotch, Welsh, and English. It does not appear from the prospectus whether the rest of the series is to be entrusted to Mr. Bonwick; but if they are it is to be hoped that he will qualify himself for the task by a preliminary study of at least the first principles of ethnology. The present volume, with all its good intentions and praiseworthy industry, must be regarded as a hopeless failure, owing entirely to the neglect of this necessary precaution. For many years ethnology, anthropology, and philology were subjects

which anyone seemed competent to deal with, who had got hold of a few lists of words in some obscure African or Polynesian dialects (the obscurer the better), or who had desecrated a sufficient number of ancient barrows, or posed to admiring circles under the shadow of some Druid's altar in Cornwall or Brittany. But those halcyon days of the amateur ethnologist are no more, though the writer, unfortunately, seems scarcely alive to the fact.

The following epigram on Dr. Siemens's recent paper has been sent us as by "a well-known scientific man". It is entitled *Electric Chlorophyll*:—

"*Quis veterum vidit plantas sine sole virentes
Germinat en semen Siementis lumine claro.*"

It appears that the Berlin Municipal Corporation has granted to Dr. W. Siemens the concession of one electrical railway which will connect Wedding-Platz with Belle Alliance-Platz. The rails will be supported by

iron columns, which will not be an obstruction for the circulation of carriages and passengers in the streets. There will be no intermediate station between the two termini.

A deplorable accident has taken place at the Grenoble Lycée. The professor of chemistry was lecturing on salts of mercury, and had by his side a glass full of a mercurial solution. In a moment of distraction he emptied it, believing he was drinking a glass of *eau sucrée*. The unfortunate lecturer died almost immediately.

Mercury was seen at Paris on May 10 and 11 with the naked eye, owing to the transparency of the atmosphere and the great elongation of the planet. It had the brightness of a 1st class star, and was of a yellowish colour. The observation was made by M.M. Henry brothers, at the Paris Observatory.

Etna is again tranquil, its summit is once more covered with snow, and an ascent is contemplated, with a view to examine the alterations caused in the crater by the recent eruptions.

From *Nature*, **21**, 18 March 464, 473-4; 1880.

Devl., Growth Differ. **21**, 175; 1979). The placement in cellular terms of the boundary between future brain and epidermis may be altered by grafting (for example Spemann & Mangold *Arch. Mikr. Anat. u. Entwmech.* **100**, 599; 1924) or by artificial imposition of physiological gradients up to this time, so that a recent claim that seven true compartments within brain structure are established within the 512-cell stage amphibian, besides being unsupported by the data produced (M. Jacobson *Trends in Neurosciences* **3**, 3; 1980), is incomprehensible and at odds with all other understanding.

There is no reason to believe that vertebrate groups other than amphibians are any more precipitate in marking out their brain plans. But most important, the means for a temporal analysis of the logic of cellular commitment, when it does take place, are denied us in these animals. This is because transition from pluripotency to quite detailed commitment within small pieces of neural tissue proceeds very rapidly in relation to the cell cycle. Cell number scarcely doubles in the neural plate and tube during the essential establishment of the brain's pattern. Marking clones will thus be of no avail in finding out whether in fact a telescoped version of the binary, sequential decision-tree that has so fascinated the *Drosophilologists*, also occurs among the cells of frog, chick and mammal embryos.

Closely associated in the insect literature with the idea of compartments is the concept of the 'selector gene'. This is based on the existence of a class of mutants called *homoeotic*, in each of which the cell populations that should belong to a particular set of the compartments instead show a tendency to construct, and thus duplicate, pattern regions proper to other compartments. Thus in one case, cells in the posterior of the anterior/posterior pairs of compartments that construct the wings (and also the obviously homologous 'halter' structures of the succeeding body segment), show a pronounced tendency to duplicate instead the structures proper to the anterior compartments, while in another case, antennal structures tend to be replaced by the leg pattern. Homoeotic mutations are now widely assumed to be in genetic units — the selector genes — whose wild-type function is literally turned on as a positive switch to define and maintain membership in one of the two compartments segregated by each restriction event at particular levels in a hierarchical sequence or 'program' (Garcia-Bellido *Am. Zool.* **17**, 613; 1977). Pairs of compartments are generated in parallel sequences from previously established ones, so that membership of each final compartment would be defined, simply and combinatorially, by the states of the set of on-off switches represented by the selector genes.

Again the idea, with its attendant image of the 'computability' of development, is

at present an attractive one. There is little doubt that the phenomena of polyclonal compartmentation and of homoeosis have revealed some sort of logical and combinatorial structure for restrictions on the potentialities of groups of cells during development. But I feel the selector gene model to be more an example of what Stent (*Trends in Neurosciences* **3**, 49; 1980) has termed the ideological component of the molecular genetic viewpoint, than anything logically required by the data. Shouldn't we do well at this stage to be flexible, rather than succumb to the current tendency for each aspect or event in an organism's life that attracts our interest promptly to become the express responsibility of a gene? Parallel manifestations of this incipient social disease in psychiatry and criminology need no comment, but they may also be seen in contemporary evolution theory (for example Dawkins *The Selfish Gene*, Oxford University Press, 1976; for a critique of the syndrome see Gould & Lewontin *Proc. Roy. Soc. B.* **205**, 581, 1979).

It was the great success of molecular biology to reveal that conservation and readout of the information that specifies proteins, and at least in prokaryotic cells, control of that readout, resembles in form the programs we design to retrieve and use information. Ideology only creeps into our view of development when it is supposed, without evidence, that the reliability and constancy of the phenomena seen there reflect an extension of the same sort of programming that controls the biosynthesis of the basic building blocks and signals. The complex and repeatable sequence of readout from the genetic library that accompanies development does not allow us to conclude that the process is actually being driven along by a digital program embodied in individual genes as molecular switches.

No experimental results yet require that any episode in the setting up of the body plan of any animal is literally the responsibility of a particular piece of DNA. The *bithorax* series of mutations in *Drosophila* (Lewis *Nature* **276**, 565; 1978) all occur in one short stretch of chromosome in a spatial order that could reflect the series of homoeotic transformations they cause between the compartments underlying the segmental plan of the insect. Here we have a hint that organisation of a stretch of genomic information may directly reflect the organisation of development, but specific selector genes need not be invoked as the targets for all or even most homoeotic mutations. Cells seem to be systems with a few stable end-states or pathways of development and a defined sequence of possible transformations between these, whether driven systematically or by chance perturbations (see Hadorn in *Major Problems in Developmental Biology* **85**, Academic Press, New York, 1967, for

other relevant phenomena in *Drosophila* cells).

Waddington (*SEB Symp.* **2**, 145; 1948) and Kauffman (*J. theoret. Biol.* **22**, 437; 1969), among others, have discussed how these observed properties might emerge at levels far removed from single genes as switches. The genome will surely contain many targets for mutations which modify various properties of all the developing cells (such as membrane transport, division rate or responses of control mechanisms to small molecules) in ways that are limiting only to their entry into certain of the available epigenetic states (that is, compartment designations). We might then expect these less available states or designations to be replaced on a systematic basis by alternative ones. Until direct molecular knowledge is at hand that justifies us in replacing views like those outlined above with the notion of a more direct programming of development written into the structure of the genes, our phenomenological knowledge about the regularities of development should be recognised as being just that. □

Continents old and new

from I. G. Gass and J. B. Wright

MANY geological texts and research papers identify two schools of thought when it comes to the subject of the origin and evolution of continental crust. One school, which could be labelled the 'conservationists' or 'reworkers', maintains that virtually all of the continental crust formed during the earliest part of the Earth's history and has been periodically reworked since. Reworking involves remelting and recycling of the original crust but not adding any new increments to it. The alternative 'accretionist' or 'incremental' hypothesis is that continental crust has grown continuously or episodically since the time of preservation of the oldest crustal rocks around 3,800 Ma ago.

A related question is the extent to which plate tectonic processes have operated during the Earth's history. At one extreme is the view that as soon as crust formed at the Earth's surface it was being moved about by plate tectonic processes; at the other extreme are those who maintain that plate tectonics did not begin to operate until late in the Proterozoic. While the meeting seemed to throw up a consensus favouring an incremental model for crustal evolution, opinion about the time of onset of plate tectonics seemed to be as divided as ever.

Several speakers provided evidence to support the thesis that (1) some 70–75% of continental crust was formed during the Archaean (> 2,500 Ma), the remaining

*A discussion meeting, 'The origin and evolution of the Earth's crust', was held in London by the Royal Society on 21–22 February, 1980 and organised by S. Moorbath and B. Windley.

25–30% during the Proterozoic and Phanerozoic; (2) crust-forming magmatic products originated in the mantle; (3) existing continental material had relatively little effect in modifying the composition of magmatic mantle products. Some took plate tectonic processes back to 3.8×10^9 years and correlated Archaean greenstones with Phanerozoic ophiolitic sequences and Archaean granitoids with products of present-day arc systems. Others pointed to geothermal and geochemical differences and argued either that if plate tectonics had existed in the Archaean, the processes were markedly different from those of the present; or that some other mechanism must have been involved.

Some interesting ideas were presented and several speakers dealt with particular aspects of accretionary processes. For instance, J.F. Dewey (Albany, New York), a staunch advocate of Precambrian plate tectonics, invoked the term 'para-oceans' and maintained that mantle processes leading to the formation of basins or para-oceans without reaching the true sea-floor spreading stage, probably make quantitatively important contributions to evolution of the continental crust. W.R. Dickinson (University of Tucson) suggested that sufficient variables exist in present-day plate tectonic processes to explain all the differences identified for older geological formations. Others, notably D.E. Karig (Cornell University), A.M. Ziegler (University of Chicago), W. Hamilton (US Geological Survey, Denver), F. Barker (USGS, Denver), R.S. Thorpe (Open University) and G.C. Brown (Open University), discussed the role of existing crust in modifying mantle products either by the incorporation of subducted sediments or by reaction between ascending magmas and the crust. Most argued that the modifications were quantitatively minor, though it was evident that a fuller understanding of island arc and back-arc processes (J. Tarney, University of Leicester) is essential. In this respect, B. Windley's (University of Leicester) report on an upended Upper Cretaceous arc in the Himalayas (traversed by the new Karakoram Highway), was particularly appropriate (see box). By way of a corrective to those advocating large scale relative movements of cratonic blocks in the Proterozoic, D.J. Dunlop (University of Toronto, Ontario) pointed out that the palaeomagnetic record remains inconclusive.

Geological, geophysical and geochemical modelling attempts to place restraints on the shape, form and composition of continental crust and hence on the processes that form it. E.R. Oxburgh (University of Cambridge) and P.R.A. Wells (Shell, The Hague) presented thermal models relating to the thickness of continental lithospheric plates and the

I.G. Gass and J.B. Wright are in the Department of Earth Sciences at the Open University.

An island arc section in the Himalayas

from Mike Coward, Geoffrey King and Brian Windley

THE newly-constructed Karakoram Highway in Pakistan provides access to (and excellent exposure of) the magnificent Kohistan island arc section. A recent conference of an International Geodynamics Working Group* provided much information on the geological and structural setting of the Kohistan island arc as well as a visit to the site.

At the conference Qasim Jan and K. Tahirkheli (University of Peshawar) and the French group from the Geology Department, Montpellier University described the structure cut by the Karakoram Highway in the Kohistan Himalayas. They interpreted this as a complete section of a late Cretaceous island arc, standing on end. At the base of the section, garnet and pyroxene rocks interpreted as mantle are overlain by amphibolite and metagabbro. Above this is a huge elongate layered metanorite lopolith at least 15 km thick and 300 km long. Above this and apparently related to it are calc-alkaline intermediate and acid intrusive and volcanic rocks topped by a magnificent section of pillow lavas and volcanic derived sediments.

This is the first complete section of continental style crust to be identified. It

*Held at the University of Peshawar, Pakistan on 19–27 November, 1979.

extent of under and overplating above destructive margins. Oxburgh argued that the continental lithosphere can be at most 250 km thick, a conclusion contested by T.H. Jordan (Scripps Institution, San Diego) who presented evidence that the lithosphere (tectosphere) beneath cratonic areas may be up to 400 km thick. However, this geologically reasonable model was strongly challenged by other geophysicists in the audience. Two contributions by H.D. Holland (Harvard University) and S.R. Taylor (Australian National University) respectively discussed the role of the atmosphere and hydrosphere and of sediments in crustal evolution; and J.V. Smith (University of Chicago) rounded off an interesting session with some speculations about the first 800 Ma of geological time (between 4,600 and 3,800 Ma ago).

The afternoon of the second day proved the most controversial part of the whole meeting. Chaired by the unashamedly 'accretionist' S. Moorbath (University of Oxford) it opened quietly enough with an exposition of the role of fluids, both magmatic and volatile, in crustal evolution (G.N. Hanson, State University of New York, Stony Brook). R.L. Armstrong (University of British Columbia) ably and entertainingly stated the 'conservationist' case, using evidence from a wide variety of sources, to support his thesis that the

is still apparently being emplaced as a massive thrust slab imbricated between the plate to the north carrying the Karakoram granite and thrust slabs of the Indian subcontinent. During the thrust process, which continued seismicity suggests is still active, the Kohistan sequence has been tilted so that the layering and bedding are vertical.

Future work on the exposed section will allow the chemical differentiation process in island arcs to be studied. Some aspects of the process are already obvious and interesting. Since the lopolith from which the higher level andesitic rocks derive is basic in composition, it seems that andesitic differentiation to form calc-alkaline rocks occurs at crustal levels rather than deeper in the mantle and the details of this can be studied.

Further studies of structure and seismicity should also elucidate the processes whereby dense material can be lifted by thrust processes to high crustal levels and may also indicate the depth at which decoupling occurs in lower lithosphere.

Mike Coward is in the Department of Earth Sciences, University of Leeds, Geoffrey King is in the Department of Geodesy & Geophysics, University of Cambridge, and Brian Windley is in the Department of Earth Sciences, University of Leicester.

continental crust has not grown significantly since early Precambrian times, but has been quantitatively recycled since it was first fully formed early in the Earth's history.

The discussion which followed showed this school of thought to be in a minority, but it also demonstrated that the difference between the two schools may be more imagined than real: few would dispute that there is some recycling of continental crust, the question seems to be simply: 'how much'. On the basis of some crustal modelling, R.K. O'Nions (University of Cambridge) argued that although most continental crust formed by about 3×10^9 years ago, the relationships between Nd, Sr, and Pb isotopes are consistent with a small amount of crustal recycling.

Finally D.P. McKenzie (University of Cambridge) focused on the geophysical constraints on models for the early Precambrian, when heat generation was much greater than now, and provided a salutary reminder that crust and lithosphere cannot be treated separately when discussing crustal evolution.

All in all, there was little if any 'crossing the floor' at this meeting, the protagonists seem to have gone their separate ways unconvinced by each other's counter-arguments. But this was a timely, entertaining and authoritative review of the contrasting lines of thought. □

REVIEW ARTICLE

Particle beam weapons— a technical assessment

G. Bekefi, B. T. Feld, J. Parmentola & K. Tsipis

Program in Science and Technology for International Security, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

This article explores the feasibility of particle beam weapons, identifies the physical principles that would govern their operation, assesses the technical difficulties, and evaluates their potential uses.

THE US and the Soviet Union are thought to be working on the design of particle beam weapon systems¹⁻³. We describe here the physics of the generation, propagation, and interaction with a target of a beam of particles intense enough to inflict damage or destroy the target. We analyse the operational aspects of a particle beam weapon system: Who could use it and for what missions. How one could track the target and aim the beam with accuracy dictated by the zero-miss requirement inherent in this type of weapon. How the damage inflicted on a target could be assessed. How the entire weapon system could be commanded and controlled, and what countermeasures could be used against it. Laser beam weapons are not considered.

We examine the possibility of an energetic beam of particles as a weapon system with either exoatmospheric or endo-atmospheric applications. Based on a satellite orbiting about the Earth, the weapon could be used to shoot down intercontinental ballistic missile (ICBM) boosters in the powered-boost phase of their orbits, after they have ascended above the atmosphere. It could also be used to attack space-borne assets of another country. And, it could attack individual reentry vehicles (RVs) as they speed towards reentry into the atmosphere shortly before they reach their intended targets. An endoatmospheric particle beam based on the ground could defend point targets such as missile silos against reentering warheads or, if placed on a ship, could defend it against cruise missiles.

For use against ICBM boosters, the particle beam weapon could be deployed on an orbiting platform either in a near-Earth orbit of ~1,000 km above the Earth's surface, or in a geosynchronous orbit, at an altitude of ~40,000 km. From the latter orbit two satellites would suffice for continuous Earth-wide coverage of all possible ICBM or submarine launched ballistic missile (SLBM) launching sites. From a 1,000 km orbit, assuming horizon-to-horizon coverage, one would require a minimum⁴ of 29 platforms with particle beam weapons. A more realistic value for the angular coverage gives⁴ 120 as the number of platforms. In all cases, to be effective as an ABM system, each weapon would have to neutralise ~1,000 missiles during the first 5 min of their flight; that is, the system must be capable of destroying one target every 0.4 s. Unlike an antiballistic missile (ABM) system that uses missiles with nuclear warheads, a particle beam weapon requires a direct hit on the target to destroy it. As the characteristic length of an ICBM booster is of the order of 10 m, and the distance to it from the platform typically 1,000 km, the beam would have to be aimed with an accuracy of better than 10 p.p.m., that is, ~1,000 times more accurately than what an ordinary ABM system requires. Also, the system would need to be designed to fire several times per second at any one of the 1,000 targets, adjusting the aim each time to compensate for the miss distance of the previous shot.

Alternately, a particle beam weapon system based on orbiting platforms could be used to attack ballistic reentry vehicles that carry nuclear warheads to US targets. This approach for ballistic missile defence has no advantages over one designed to attack

the missile boosters. Reentry vehicles are considerably smaller than an ICBM booster, which complicates the problem of their detection, identification, and the aiming accuracy. Although they may be targetable from an orbiting platform for more than a few minutes, their large numbers (the Soviet Union is deploying missiles with 8-10 RVs per booster) require faster system response.

A third exoatmospheric use of a particle beam weapon would be as an anti-satellite system. When used for long-range applications against satellites in very high orbits above the Earth, the system would encounter the same problem of beam aiming and tracking as in the ABM use, except that in this case the target could be held under attack for a long period of time.

Endoatmospheric applications would include ground- or ship-based systems either for the point defence of silos against reentry vehicles, or of ships against cruise missiles. An ICBM RV on a normal trajectory would rise above the radar search horizon some 4,000 km, or ~10 min travel, away from the radar. For a cruise missile travelling near the surface at a speed of Mach 1, the radar detection distance could be as short as 1 km or about 4 s away.

To determine the performance parameters and physical characteristics of a particle beam weapon we began at the target, by first calculating the amount of energy that would damage a target, then examined the propagation of a particle beam through the atmosphere or the vacuum of outer space, and finally specified the characteristics the beam must have at the exit of the accelerator. On the basis of such considerations, one can calculate the required parameters of the accelerator.

Our assessments sometimes use idealised models of, often, very complex physical phenomena. However, our calculations tend to overestimate the effectiveness of the beam as a weapon.

Damage mechanisms and energy deposition on target

To determine the energy per unit area that the particle beam must deposit to be an effective weapon, one must consider the mechanisms by which deposition of this energy on a target would destroy it or cause it to malfunction. The mechanisms considered are: boring a hole in the outer shell of the ICBM or cruise missile, melting or damaging the guidance and/or control electronics of any of the targets; or exploding the high explosive trigger(s) of the nuclear warhead(s) carried by an ICBM or cruise missile.

A charged particle can cause chemical explosives to detonate by several cooperative mechanisms. Most chemical explosives will explode if their temperature is raised to ~500 °C when confined. In the presence of the beam, this threshold temperature may well be lower, because of electric field and shock wave creation within the material. Thus, $\Delta T = 500$ °C is a conservative figure. The energy density ε required to detonate the explosive is then given by $\varepsilon = 3R\rho\Delta T/M$ where R is the universal gas constant, and ρ and M are the mass density and

molecular weight of the explosive, respectively. Taking $\rho = 0.8 \text{ g cm}^{-3}$ and $M = 50$, yields $\epsilon = 250 \text{ J cm}^{-3}$. Similarly, knowing the specific heat, melting temperature and the heat of fusion of materials, the energy density necessary to melt them can be computed. To melt the aluminium shell of the weapon requires $\epsilon \approx 3,200 \text{ J cm}^{-3}$ and to melt the silicon substratum of electronic components requires $\epsilon \approx 7,000 \text{ J cm}^{-3}$.

Electronics, however, are not only sensitive to catastrophic failure due to an increase in temperature. Rate-dependent effects, caused by the creation of a large number of electron-hole pairs generated during the passage of a charged particle beam through the electronics, can also destroy the circuits. For example, recombination of the pairs could generate enough heat to damage circuit elements. The threshold for this effect in silicon is $\sim 10^{13} \text{ rad s}^{-1}$ or $2.4 \times 10^8 \text{ J cm}^{-3} \text{ s}^{-1}$. Thus, a beam that could deliver 10^3 J cm^{-3} in $1 \mu\text{s}$ could damage electronic circuits via this mechanism. Alternatively, production of ions which are trapped inside the microcircuits can form inversion layers which interrupt operation. Incident radiation can shift switching thresholds and cause source leakage in metallic oxide silicon circuits, which drains the power supply. Switching times can be affected: an energy deposition of $\sim 25 \text{ J cm}^{-3}$ would be enough to cause havoc in the guidance or control electronics of a missile or satellite. These results are summarised in Table 1.

Table 1 Estimated minimum deposition energy densities on various targets

Material	Type of damage	Energy density ϵ (J cm^{-3})
Aluminium	Melting	3,200
Chemical explosive	Detonation	250
Silicon	Melting	7,000
Silicon	Electron-hole recombination in circuits	1,000
Semiconductors	Shifts in circuit switching thresholds and times	25

The total energy E that must be deposited on target to cause damage is given by $E = \epsilon l A$ where ϵ is the appropriate energy density as given in Table 1; $l = E \text{ dz/dE}$ is the scale length for energy absorption, that is the distance over which the beam particles are degraded in energy to $(1/e)$ th of their initial value, as a result of passage through the medium; A is the cross sectional area of the beam illuminating the target, and depends on the optics of the beam but not on target size.

A beam for a functional weapon seems to be one with $\epsilon \approx 1,000 \text{ J cm}^{-3}$ and $A = 10^4 \text{ cm}^2$. With $l \approx 9 \text{ cm}$, a value appropriate to the passage of an electron beam of several hundred MeV through aluminium, one obtains $E \approx 100 \text{ MJ}$. This then, is the energy that must be supplied by the accelerator. If V is the beam voltage, I the beam current, and τ the pulse length, we have that $E = V I \tau = 10^8 \text{ J}$. Setting $V = 10^9 \text{ V}$ and τ , say, equal to $100 \mu\text{s}$, gives $I = 1,000 \text{ A}$. Note that if the beam is to have a cross-sectional area $A = 10^4 \text{ cm}^2$ on a target $1,000 \text{ km}$ away, the required angular beam divergence is only 10^{-6} rad , which is a very demanding accelerator beam emittance.

Exoatmospheric propagation of a particle beam

We now examine three effects which impede the propagation of the charged particles in the vacuum of outer space, and then consider systems using neutral beams.

Electron and proton beams: The first effect is the radial spreading of a charged particle beam caused by Coulomb repulsion between like charges. Using relativistic particle dynamics, one finds that the beam radius a at a distance z from the source is

determined from the expression,

$$aF(\ln(a/a_0)^{1/2}) \approx (z/\beta\gamma)(I/I_A)^{1/2} \quad (1)$$

where a_0 is the initial beam radius and F is Dawson's integral⁵; $\beta = v/c$ with v as the axial particle velocity, and $\gamma = [1 - (v/c)^2]^{-1/2}$. I is beam current and I_A the Alfvén current defined as⁶

$$\left. \begin{aligned} I_A &= 4\pi\epsilon_0 m_0 c^3 \beta\gamma / e \\ &= 1.70 \times 10^4 \beta\gamma \text{ A for electrons} \\ &= 3.12 \times 10^7 \beta\gamma \text{ A for protons.} \end{aligned} \right\} \quad (2)$$

For a $1,000 \text{ A}$, 1 GeV electron beam having an initial radius of 1 cm , and striking a target at a distance $z = 10^3 \text{ km}$, equation (1) yields $a \approx 15 \text{ m}$. For an identical proton beam, $a \approx 1.8 \times 10^4 \text{ m}$. Therefore, while electron beams cannot be entirely ruled out as a weapon by this phenomenon alone, proton beams would obviously have no practical use.

The second effect which interferes with the use of charged particle beams in outer space is the deflection they suffer in the Earth's dipole magnetic field. In a uniform magnetic field $\mathbf{B}$, the radius of curvature⁷ $R = (m_0\beta c/eB)\gamma$. Hence a 1-GeV electron beam directed from a $1,000 \text{ km}$ orbit above the Earth towards a target $100\text{--}200 \text{ km}$ above the Earth's surface will have a radius of curvature in the geomagnetic field of $\sim 110 \text{ km}$. A corresponding proton beam will have a 190 km radius of curvature. Therefore, these beams will never reach the target $1,000 \text{ km}$ away. We show below that higher energies will be of no use.

Charged particle beams are deflected from their original direction by the geomagnetic field. The uncertainty ΔR in the position of a 1-GeV electron beam due to an uncertainty of 1% of the value of the geomagnetic field is 100 m at a $1,000 \text{ km}$ range, as obtained from the equation $\Delta R = R(\Delta B/B)$. This forbids the aiming of the beam. Another serious problem is that the geomagnetic field can be greatly and unpredictably disturbed by a nuclear explosion in the stratosphere.

The alternative is to use neutral particle beams for exo-atmospheric applications. We have considered neutron, γ -ray, and neutral hydrogen beams. The first two are unsuitable because the beam divergence is intolerably large. The neutral hydrogen beam is more promising.

Neutral hydrogen, H^0 , beams: A neutral hydrogen beam is generated from an H^- beam, that is, hydrogen atoms with one extra electron. Beams of these particles can be accelerated⁸ to $\sim 500\text{-MeV}$ energies in circular accelerators with a confining magnetic field of $\sim 10,000 \text{ G}$. The maximum energy achievable by this process is limited by the Lorentz force. That is, the particle feels an electric field of magnitude $q[\mathbf{v} \times \mathbf{B}]$ which causes distortion of its atomic potential and, if sufficiently large, leads to subsequent breakup of the atom. In a linear accelerator, because of relativistic effects, it is difficult to achieve energies $\geq 800 \text{ MeV}$.

The beam of H^- particles is passed through a system of magnetic lenses to focus and aim it. It is then neutralised by removing the extra electron by various methods such as passing it through a gas stripper, the Lorentz force generated by an externally imposed magnetic field, a plasma cell, or by ejecting one of the bound electrons with a laser beam. The gas-stripper approach seems to be the most direct. The cross-section for the stripping reaction is large (typically one-tenth of the geometric cross-section of the atom) and the efficiency^{9,10} of converting H^- to H^0 can be as high as 50% after passage through a few microgrammes of gas. To estimate the minimum divergence θ of the H^0 beam, we assume that the incident H^- beam has zero transverse momentum when incident on the gas stripper. From the uncertainty principle it is readily shown that $\theta \approx 2[(m_0/M)E_{H^-}/E]^{1/2}$ where m_0 is the electron mass, M the proton mass, E_{H^-} the binding energy ($\sim 1 \text{ eV}$) of the electron in the H^- ion, and E the energy of the beam. Setting $E = 500 \text{ MeV}$, gives $\theta \approx 2 \times 10^{-6} \text{ rad}$, a value which is sufficiently small to make the neutral hydrogen beam of potential weapon interest.

There are, however, several problems. First, the beam divergence calculated above represents a lower limit as it ignores the divergence of the incident H^- beam; the total angular spread due to the addition of spreads from two (independent) sources is given by $(\theta_0^2 + \theta^2)^{1/2}$ where θ_0 is the spread of the incident H^- beam. Therefore, it is not certain that accelerator technology will make it possible to construct an accelerator having the necessary small emittance, large beam current, and large beam energy. Second, if the beam size is to be maintained, the magnetic lens that bends the beam must be achromatic to 1% or better and have a stable and predictable field strength to 1 p.p.m., requiring that the system of magnets be carefully shielded from the geomagnetic field. And third, the determination of the miss vector is very difficult with a neutral beam.

One can also generate a neutral beam through the capture of electrons by a beam of positive ions, leading to the formation of "plasmoids". However, such beams have large beam divergencies and are not suitable for weapon use.

Endoatmospheric propagation of a particle beam

When a charged particle beam is fired into the atmosphere, it partially ionises the gas in its immediate neighbourhood. The plasma thus generated helps to reduce the radial electric field associated with the primary beam, to the point that it is balanced by the beam's self-magnetic field. In principle, then, it is possible to propagate a charged particle beam in the atmosphere. In the following we examine its propagation characteristics.

Beam propagation: The magnitude of the allowed beam current I is constrained by the fact that as the current in, say, a cylindrical beam increases, so does its self-generated azimuthal magnetic field B_θ . An individual beam electron subjected to this field will undergo radial oscillations, and if I and thus B_θ are large enough, it will be deflected backwards⁶, opposite to its initial (axial) propagation direction. When the beam is completely space charge neutralised by the presence of the ambient ionised gas, and if moreover the beam has a uniform density over its cross-sectional area, one then finds⁶ that the maximum allowable current $I_{\max}$ is the Alfvén current I_A given by equation (2). However, by proper beam shaping^{11,12} (for example, using a hollow, annular beam) $I_{\max}$ can substantially exceed I_A . When a uniform beam is partially space charge neutralised, with a neutralisation factor f ($0 \leq f \leq 1$), the appropriate maximum current is now given by¹³

$$I_{\max} = 17,000 \beta^3 \gamma / (\beta^2 + f - 1) \text{ A} \quad (\text{for electrons}) \quad (3)$$

Equations (2) and (3) refer to infinitely long beams. However, one needs to consider the beam front where a rising current occurs. The time varying current generates a time varying azimuthal magnetic field $B_\theta(t)$ which in turn produces a time varying electric field $E_z(t)$ that drives a counterstreaming current in the ambient plasma. This is referred to as 'current' or 'magnetic neutralisation'. When magnetic neutralisation is complete (that is, the forward primary beam current equals the induced counterstreaming current) $I_{\max}$ is unlimited in magnitude. If, however, magnetic neutralisation is only partial, one finds that,

$$I_{\max} = 17,000 \beta \gamma / (1 - f_m) \text{ A} \quad (\text{for } f = 1) \quad (4)$$

where f_m is the fractional magnetic neutralisation factor, $0 < f_m < 1$.

Knowledge of the magnitudes of f and f_m and of beam propagation characteristics in general requires detailed knowledge of ionisation cross-sections of atmospheric constituents and of beam dynamics. Partial (or complete) space-charge neutralisation f comes about through two processes: by direct ionisation of the gas by the beam, and by avalanching in the induced electric field¹⁴. In addition to these 'classical' processes, and at pressures below ~ 1 torr, ionisation by microinstabilities is prevalent. Here electron generation occurs as a result of the

strong turbulent electric field associated with the instabilities^{15,16}.

A typical experimental value of the ratio n_p/n_b of the secondary plasma density to primary beam density is, $10^3 \leq n_p/n_b \leq 10^4$. For example, when the beam current density $J_b = 1 \text{ kA cm}^{-2}$, $n_b = J_b/ec = 2 \times 10^{11} \text{ cm}^{-3}$ and it follows that $2 \times 10^{14} \leq n_p \leq 2 \times 10^{15} \text{ cm}^{-3}$. As relevant gas pressures lie in the range 4×10^{16} to $3 \times 10^{19} \text{ atom cm}^{-3}$ (atmospheric pressure), we see that the ambient gas becomes weakly ionised (degree of ionisation $\leq 2\%$), and the beam forms an electrically conducting channel of conductivity σ , where

$$\sigma = n_p e^2 / m \nu \quad (\text{S m}^{-1}) \quad (5)$$

Here ν is the electron-atom collision frequency for momentum transfer (for air¹⁷ $\nu \approx 1.8 \times 10^{-7} n_g \text{ s}^{-1}$ where n_g is the gas atom density in cm^{-3}). At $p = 2$ Torr, $n_g = 7 \times 10^{16} \text{ atom cm}^{-3}$, and setting $n_p = 2 \times 10^{14} \text{ cm}^{-3}$, yields $\sigma = 460 \text{ S m}^{-1}$.

Magnetic neutralisation f_m becomes possible¹⁸ when the conductivity σ is large enough, enabling charge of the ambient plasma to move in the axial electric field E_z generated by the propagating electron beam. This electric field drives the return current I_r , whose current density is $J_r = \sigma E_z$. The magnetic neutralisation factor $f_m \equiv J_r/J_b$ is then given by $f_m = \sigma E_z/J_b \approx \sigma V/J_b L$ where V is the axial voltage drop along the beam, and L is its length. Setting $f_m = 1$, $\sigma = 460 \text{ S m}^{-1}$ and $J_b = 10^7 \text{ A m}^{-2}$, yields an axial electric field $E_z = 2.2 \times 10^4 \text{ V m}^{-1}$. We observe that the axial electric field causes an energy degradation of the primary beam. In going 1 km, each beam electron loses 22 MeV, a 2% energy loss from a 1 GeV beam. This, of course, is not the only energy loss¹⁹ by the beam particles.

Table 2 summarises qualitatively the propagation characteristics¹⁸ of an electron beam. Much experimental and theoretical work is needed for a clear understanding of this very complex problem.

Table 2 Propagation characteristics¹⁸ of an electron beam in air*

Air pressure (torr)	Typical value of f	Typical value of f_m	Beam behaviour
$< 10^{-3}$	≈ 0	≈ 0	Beam blows up
$\approx 10^{-1}$	≈ 1	$\ll 1$	Beam pinches
≈ 1	1	≈ 1	Beam drifts in a force-free environment
> 20	1	≈ 0	Beam pinches and then breaks up

* $V \approx 3 \text{ MeV}$, $I_b \approx 50 \text{ kA}$, pulse length $\approx 30 \text{ ns}$

Beam dispersion: Consider first an electron beam. To determine the increase in area along the beam axis caused by scattering, we assume that the beam is fully space-charge neutralised and that the ratio of beam current, I_b , to the Alfvén current, I_A , is $\ll 1$. This corresponds to the paraxial ray approximation for which the transverse particle velocity is small compared with the longitudinal velocity. The beam current density and the fields are taken to be azimuthally symmetric, with no density variations occurring across a transverse segment of the beam. The energy of the beam particles is assumed to remain constant, independent of distance. Only small-angle multiple scattering is considered, beam segments do not overtake one another in the direction of propagation, and effects of return current are neglected. Particles interact with each other by the collective fields they induce. All relevant time scales are long compared with the period of particle oscillation. In these conditions one obtains a form of Nordsieck's equation^{20,21} given by

$$(I_b/I_A) \ln [R^2(z)/R_0^2] = \langle \theta_s^2(z) \rangle - \langle \theta_0^2 \rangle + O(d^2 R^2/dz^2) \quad (6)$$

where I_b is the beam current and I_A the Alfvén current given by equation (2). $R(z)$ is the beam radius at some axial distance z , and R_0 is the beam radius at the injection point $z = 0$. The quantity $\langle \theta_s^2(z) \rangle$ is the mean square scattering angle due to scattering by the background gas and $\langle \theta_0^2 \rangle$ is the mean square

angle of the random component of the particle motion at the point of injection, $z = 0$. It is related to the accelerator emittance ε through²⁰

$$\langle \theta_0^2 \rangle = (\varepsilon/R_0)^2 \approx (I_b/I_A) \ll 1 \quad (7)$$

For small beam spreading, the term $O(d^2R^2/dz^2)$ appearing in equation (6) can be neglected.

The mean square scattering angle is given by²²

$$\left. \begin{aligned} \delta \langle \theta_s^2 \rangle / \delta z &= c_1 \ln(c_2 z^{1/2}) \\ c_1 &= (8\pi e^4 Z(Z+1)n) / (m^2 c^4 \gamma^2 \beta^4) \text{ cm}^{-1} \\ c_2 &= (hZ^{2/3} n^{1/2}) / (\pi^{1/2} mc\beta) \text{ cm}^{-1/2} \end{aligned} \right\} \quad (8)$$

where Z is the atomic number of the gas (equal to ~ 7.4 for air) and n is the density of nuclei (equal to $2.69 \times 10^{19} \text{ cm}^{-3}$ at sea level). Integrating equation (8) with respect to z , and inserting the result in equation (6) yields

$$R^2(z)/R_0^2 = A(z)/A_0 = \exp [I_A/I_b] (c_1 z \{ \ln [c_2 z^{1/2}] - 1/2 \}) \quad (9)$$

where $A(z)$ is the cross-sectional area of the beam at position z and A_0 is its area at the accelerator, $z = 0$.

The larger the beam current (consistent with $I_b/I_A \ll 1$) the smaller the beam dispersion. We take as an example our model beam: beam voltage = 1 GeV, beam current = 1,000 A. Then, at a distance z of 1 km, equation (9) gives $A(z = 1 \text{ km})/A_0 = 7.0 \times 10^7$. A beam which is initially, say, 1 cm in radius expands to 84 m in a distance of 1 km. Note, however, that this is an optimistic lower limit on the dispersion suffered by the beam. For beams with energies $\geq 100 \text{ MeV}$, the principal energy loss¹⁹ is through bremsstrahlung. The beam energy degrades rapidly as a result of this loss, and if E_0 is its initial energy, then at a distance z , $E(z) = E_0 \exp(-z/z_0)$ where for air the 'radiation length' $z_0 \approx 300 \text{ m}$. As a result, the scattering increases since in accordance with equation (8), $\delta \langle \theta_s^2 \rangle / \delta z$ varies as $[E(z)]^{-2}$. Thus, even in a distance z as short as one radiation length, the electron beam disperses hopelessly.

Now consider a proton beam with the same specifications of energy and current (1 GeV, 1,000 A). Blanchard²³ obtains for the mean scattering for heavy particle incident on matter,

$$\langle \theta_s^2 \rangle = \langle \theta_0^2 \rangle + 2\kappa Z(m_e/m_p) \ln \left[\frac{E_0(E + 2m_p c^2)}{E(E_0 + 2m_p c^2)} \right] \quad [10]$$

where E_0 and E are the proton kinetic energies at the beginning and the end of the path, respectively, and κ is a coefficient of the order of 1 (≈ 0.75 for air), having a weak dependence on the nature of the ambient medium and on the proton energy. Using the results of Bethe and Ashkin²⁴ for proton energy loss due to both hard and soft collisions, one finds, for protons with an initial energy $E_0 = 1 \text{ GeV}$, a loss of $\sim 130 \text{ MeV}$ in traversing 1 km of atmosphere. Hence $E = 870 \text{ MeV}$ and equation (10) yields $\langle \theta_s^2 \rangle \approx \langle \theta_0^2 \rangle + (5.5 \times 10^{-4})$. Substituting this result in equation (6) and noting that for protons $I_A = 3.12 \times 10^7 \beta \gamma A$ (see equation (2)), it follows that $A(z = 1 \text{ km})/A_0 \approx 10^{13}$. Beam dispersion is immense. Indeed, A/A_0 is so large that equation (6) from which this ratio was derived has long ceased to be valid. What actually happens is that the protons do not propagate as a beam but form a cloud of positive charge near the outlet of the accelerator. In addition, above 1 GeV, endoatmospheric propagation of protons is limited primarily by nuclear collision which results in a particle number loss from the beam.

Hole-boring: The above results indicate that electron or proton beams are not suitable for endoatmospheric transport of energy over any distances useful for a weapon. An alternate approach is to attempt to bore an evacuated channel through the intervening atmosphere to the target through which a particle beam could propagate. This reduces the density n of scatterers and thus the scattering angle θ_s (see equation (8)).

To calculate the amount of energy that will be removed from the beam to bore a hole, consider a pulsed 1-kA electron beam of 1-GeV energy. The particles will lose energy by two effects: ionisation and bremsstrahlung. For air at STP, the ionisation loss

is²⁵ $\sim 2 \times 10^3 \text{ eV cm}^{-1}$ for energies of $\sim 1 \text{ GeV}$. The ratio of radiation loss to collisional loss is determined by the Z of the scattering material and the energy of the beam, and is roughly $EZ/1600 m_0 c^2 \approx 9.1$.

Take a beam with an area $A_0 = 1 \text{ cm}^2$ at the accelerator port and an initial current I_b of 1.0 kA. The beam will then deposit energy by ionisation at a rate of $2 \times 10^6 \text{ W cm}^{-3}$ and by bremsstrahlung (mainly in the forward direction) at a rate of $1.8 \times 10^7 \text{ W cm}^{-3}$. Both of these effects will contribute to superheating of the air in the region of the beam, and the consequent formation of a channel of reduced density through which successive pulses could pass. This is the concept of hole-boring. To reduce the gas density in the channel to, say, $0.1 \times$ atmospheric density, the temperature must rise to $\sim 3,000^\circ \text{C}$. For a channel of 1 cm^2 cross-sectional area and 1 km long, the beam must deposit an energy of $\sim 1.5 \times 10^6 \text{ J}$.

By reducing the air density in the channel from atmospheric to $0.1 \times$ atmospheric, the radiation length for bremsstrahlung loss by electrons increases from ~ 300 to $\sim 3,000 \text{ m}$. Thus, in propagating 1 km, energy loss can be neglected, and γ of equation (8) remains essentially constant. Beam dispersion is calculated in the same way as was done in the previous subsection. For a 1-GeV, 1-kA electron beam one finds that $A/A_0 = 4.4$, which is an acceptably small spreading for weapon use.

Hole boring is indeed necessary for stable beam propagation. Table 2 and experiments^{14,26} indicate that there is a range of pressures within which stable electron beam propagation may exist. For air at room temperature the 'pressure window' extends from ~ 1 to 10 torr. The good propagation characteristics are interpreted²⁷ as a combination of two effects: the quenching of a two-stream instability by sufficiently high plasma density and electron-atom collisions; and the simultaneous suppression of the resistive hose instability caused by sufficiently high electrical conductivity generated by the ionisation processes described earlier.

Endoatmospheric electron beam propagation might, therefore, be an interesting candidate as a weapon, provided that a hole can be bored through the atmosphere that would have a neutral density n around 0.3 to $3 \times 10^{17} \text{ cm}^{-3}$. The crucial question is will the beam remain in the channel over distances of, say, 1 km? Reported experimental work^{14,26} on this problem involves laboratory measurements over propagation distances less than several metres, and with beams having energies of the order of 10 MeV. Little is known about electron beams with energies in the GeV region.

Accelerator characteristics

There are two types of accelerators in existence: (1) conventional high voltage, low current machines capable of delivering high quality beams of modest energy per pulse; and (2) accelerators based on novel technology^{28,29} developed during the 1960s. The latter are characterised by relatively low voltages, but high currents and exceedingly large beam energies. The properties of the two types of accelerators are summarised in Table 3, together with a list of parameters desirable in an accelerator meant for a particle beam weapons system: the parameters of existing machines are far removed from the needs of a weapon's type of generator.

An exoatmospheric particle beam weapon requires a machine that would accelerate a beam of H^- atoms to about 1,000 MeV. H^- ion beams are routinely accelerated in proton linear accelerator systems at the Argonne National Laboratory and at the Los Alamos Meson Physics Facility. Recent developments in ion source technology, both in the US and the USSR⁸, indicate that high-current/low-emittance H^- atom beams are possible, with currents of $\sim 0.15 \text{ A}$, and a beam divergence of $\sim 20 \mu\text{rad}$. As the current is at least three orders of magnitude less than that of what a beam weapon would require, technological developments in the near future are unlikely to achieve the requisite currents ($\sim 1 \text{ kA}$) and beam divergence ($\sim 1 \mu\text{rad}$) for a beam weapon accelerator.

The electrostatic acceleration of the beam to adequate energy for injection into a r.f. linac is done by several different accelerator types, such as a van de Graaf, and offers no technological problem. The main section of the r.f. linac, however, is a problem. None of the existing machines has all of the properties necessary for exoatmospheric particle beam weapon applications. For instance, the BNL and the Fermi Lab injector linacs both accelerate 0.15–2 A currents in short pulses. Moreover, existing proton linac accelerating gradients are low, $\sim 2 \text{ MeV m}^{-1}$. Experience with electron linear accelerators indicates that gradients of up to $\sim 30 \text{ MeV m}^{-1}$ may be sustained. This makes for extremely long machines when $\sim 1 \text{ GeV}$ beams are desired.

Table 3 Characteristics of the two types of large accelerators, and the desired characteristics of an accelerator for use with particle weapons

Parameter	High-voltage, low-current accelerator	Low-voltage, high-current accelerator	Particle-beam weapon accelerator
Voltage (MeV)	1500	15	1,000
Current (A)	0.025	10^5	1,000
Pulse length (μs)	1.5	0.12	100
Energy per pulse (J)	56	1.8×10^5	10^8
Pulse repetition rate	50 pps	Several per day	≥ 100 pps
Beam divergence (rad)	$\geq 10^{-5}$	$\geq 10^{-2}$	$\sim 10^{-6}$ *

* The quoted value of $1 \mu\text{rad}$ is for exoatmospheric weapons systems only; for endoatmospheric use, much poorer emittances would suffice.

Providing adequate power for an accelerator in space should not pose serious technological problems. The detonation of explosives coupled into appropriate generators could provide the power required. A device for staging the energy to the time scale required for an accelerator, with the capacity to store enough energy with easy retrieval at high efficiency and in a short time, would be necessary. Such staging devices are not apparently beyond current technological capabilities. If we assume an (optimistic) energy conversion efficiency from staging system to beam energy via the accelerator of 50%, and an energy conversion efficiency from primary power to the staging system of 30%, the primary power required will be about six times more than the beam power. Therefore, an accelerator that produces a 100-MJ beam (see Table 3) will require 600 MJ of primary energy per pulse. The energy stored in high explosives is $\sim 4 \times 10^3 \text{ J g}^{-1}$, and it follows that the requisite primary energy can be supplied by the explosion of $\sim 150 \text{ kg}$ of high explosives per pulse.

Another type of accelerator, the linear induction accelerator, is capable of short pulses of very high current and has been developed for various applications. A typical linear induction accelerator is the ERA 4-MeV induction linac built in Berkeley. The accelerator produced a current of 1,200 A for 40 ns at an output energy of 1–4 MeV. The system is capable of reproducible high current operation; it is, however, bulky and heavy as it consists of iron or ferrite cores for the transformer coupling. In the Berkeley ERA injector case the energy gradient was 0.73 MeV m^{-1} . Presumably this gradient can be increased to about twice this value, but a 1-GeV accelerator will be at least 400–500 m long, with current technology. For a 1,000 section accelerator the weight of iron required is about 100 tons, a reasonable value. The electron gun system necessary for this accelerator is available, and the emittance is suitable for a small diameter output beam.

The unresolved technical questions are whether the energy gradient in such machines can be increased to the point that the overall size of the accelerator becomes practical, and whether a

few kilohertz pulse rate that seems essential for useful operations can be achieved.

Operational requirements of a particle beam weapon system

The operational requirements that various missions dictate are:

- Detect and track the target(s) with appropriate sensors.
- Identify the target(s) among decoys.
- Point the beam at a target and follow the target.
- Fire the beam towards the target.
- Determine if the target was hit, or not.
- Assess the damage if the target was hit.
- Determine the miss vector if the target was not hit.
- Correct the aiming of the beam by the miss amount.
- Fire the accelerator again.
- Determine the miss distance or assess the damage to the target once again.
- Repeat this cycle until the target is verifiably destroyed.
- Communicate the results to a command center.
- Repeat the cycle with the newly engaged next target.

We consider below the performance requirements of the system in several conditions.

Antisatellite, space-based system: In this case the target object would have to be detected by long-range radar and its position further refined by means of an optical sensor. Since the target is typically a few metres in diameter, the weapon system would require optics with at least a 50-cm mirror to resolve the object 1,000 km away. If the target object is $\sim 40,000 \text{ km}$ away, the mirror would have to be 20 m. Even then, the weapon system could not readily assess either the damage inflicted on the target or the miss vector if the target were not hit, and therefore the weapon's operator has no way of knowing whether or where to re-aim the beam for a second burst. He has two choices: either fire once in a 'blind' fashion and assume that the target is impaired, or fire repeatedly until the target satellite tumbles or explodes.

Antiballistic, space-based system: A space-based boost-phase ballistic missile defence particle beam system would have to have similar subsystems as above for target detection and tracking, beam aiming and damage assessment, but the time constants of this operation are very different from an A-SAT mission. A single particle beam weapon must be capable of detecting, identifying, and tracking as many as 1,000 boosters while still in their boost phase, that is, for at most 400 s. The tracking and aiming mechanism must track each booster with an accuracy of one part in 10^5 . The aiming mechanism for a neutral beam could consist of a precalibrated magnetic lens and of focusing magnets which can be mechanically steered for accurate aiming by means of an optical telescope. This would require an achromaticity of the magnetic system of 10^{-6} so that the beam divergence could be maintained at a minimum. As the variations of the geomagnetic field are not known to such accuracy, the entire magnet complex would have to be magnetically shielded. It is very difficult to ascertain the size of the miss vector between a neutral beam and the target, and it may be necessary to fire repeatedly but blindly at a target until there are visible catastrophic results: the ICBM either explodes or tumbles out of control. An alternative to this lengthy process is to preprogram the weapon to fire stochastically a fixed number of particle pulses at a target and then shift to another one without positive identification of a successful hit.

One may also contemplate tracking the H^0 beam with a laser that would induce resonance fluorescence³⁰ (resonant Rayleigh scattering) in the hydrogen atoms. This would necessitate the use of an as yet unavailable, pulsed, tuneable laser emitting $\sim 1 \text{ J}$ of energy in the $1,000 \text{ \AA}$ wavelength range. The number n of fluorescent photons backscattered into the receiving-transmitting mirror of area A situated at a distance R from the scattering volume V is given by

$$n \approx \left(\frac{E}{\hbar\omega} \right) \left(\frac{A}{A_s} \right) \left(\frac{r_0}{R} \right)^2 \left(\frac{\omega}{\Delta\omega} \right)^2 N_H V \quad (11)$$

Here E is the laser energy and ω is its frequency; $\Delta\omega$ is the line width of the laser, or of the resonance H^0 transition, whichever is the larger; $r_0 = e^2/4\pi\epsilon_0 m_0 c^2$ is the classical electron radius, A_s the cross-sectional area of the scattering volume V , and N_H is the number of density of H^0 atoms. Taking $E = 1\text{ J}$, $\hbar\omega = 10\text{ eV}$ ($\lambda = 1,240\text{ \AA}$), $R = 10^3\text{ km}$, $V = 100\text{ m}^3$, $A/A_s = 0.018$, $\omega/\Delta\omega = 1.1 \times 10^7$, and $N_H = 2.1 \times 10^{13}\text{ m}^{-3}$ (which corresponds to a 1 kA beam of 1 m^2 cross-sectional area), yields $n = 2.3 \times 10^4$, which suffices for detection purposes. Note that in obtaining this value, we chose a diffraction limited beam divergence of 10^{-6} rad and an (optimistic) line width $\Delta\omega = 2\omega^2 r_0/3c$ equal to natural line broadening.

To track the H^0 beam and determine its position requires the laser to have a much greater pulse repetition rate than that (≥ 100 pulses per s) of the hydrogen beam. Knowing the direction of the beam axis does not resolve the problem of target tracking and aiming discussed above.

We conclude that even if an H^- atom accelerator and stripper system that would produce a neutral beam suitable as a lethal mechanism could be built, it would still be technically infeasible to deploy a particle beam weapon system that could be even marginally effective against a massive ICBM attack on the US. It is unreasonable to assume that a boost-phase ballistic missile defence system would operate in a cooperative environment, as it would have to operate mostly within line of sight of enemy territory. We must consider the countermeasures that a military commander could take against the threat of a particle beam space-borne ballistic missile system. The most obvious method would be to destroy or mine the particle beam accelerator. Weapons based in space during peace time are very vulnerable to small explosive charges on small satellites which have been placed in orbit very close to them.

The operations of a space-based particle beam weapon system would probably be closely controlled from the ground. Interfering with the link between the weapon platform and its ground control would effectively nullify the weapon. Another countermeasure is to deny the system knowledge of the precise location of its targets. This could be done either by jamming the radar which the system must use for initial detection and localisation of its targets, or by the use of offset decoys for radar, visible, or IR. For space targets one can use a chaff shroud or umbrella so that even if the particle beam system could detect the general location of the target by radar, it could not see through the shroud. As the chaff or thin aluminium sheets can be carried by the ICBM booster, such countermeasures would be effective in the case of sea-launched as well as land-based ICBMs.

Another effective countermeasure against a neutral hydrogen atom beam weapon based in space is to interpose a thin layer of air between it and the ICBMs that it is designed to attack during their boost phase. The atoms in the neutral hydrogen beam will be stripped of their electrons and become protons. As indicated above, protons of that energy cannot propagate in space as a beam because they disperse rapidly. The stripping cross-section of a 1-MeV hydrogen atom incident on a nitrogen molecules is $\sim 7 \times 10^{-17}\text{ cm}^2$ (refs. 31, 32). The cross-section falls off roughly logarithmically with increasing energy, and we estimate that at 1-GeV the cross-section is $\sim 10^{-17}\text{ cm}^2$. Now consider filling a volume $4 \times 10^{24}\text{ cm}^3$, an area of $2,000 \times 2,000\text{ km}$ and $1,000\text{ km}$ high, with sufficient air molecules to cause stripping. Using the above value of the cross-section, one finds that the mass of air required is $5.5 \times 10^5\text{ tons}$, and the energy necessary to lift it to the requisite height is $2.7 \times 10^{15}\text{ J}$. As 1 g of high explosives has $\sim 4 \times 10^3\text{ J}$, an explosion of a 0.7 Mton nuclear warhead could provide the energy necessary to raise the layer of air that would disperse the particle beam.

Ship-based, counter-cruise missile system: This mission includes the same operations outlined for the exoatmospheric weapon, except that the command and control systems will be integrated into the ship-based weapon and therefore the communications link cannot be interfered with. Another difference is that the counter-cruise missile weapon has only $\sim 3\text{ s}$ to engage and

destroy its target if it starts its attack at 1 km range, and that the number of targets at any instant would be fewer than in the ABM case, perhaps no more than three or four.

If the weapon does not shoot down the missile with the first beam burst, it must track the target and shoot again. It is not clear, however, that the beam can track a target that has any significant angular velocity with respect to the ship. To accomplish that, the beam would have to leave the evacuated channel that it propagated in and bore another very near the previous one. It is not known whether the atmosphere can support stably such a procedure or how this procedure depends on the angular velocity of the target, the size of the channel, the state of the atmosphere in the immediate vicinity of the ship, or how reproducible such behavior could be.

If the attacking cruise missile can manoeuvre in response to an attack by a charged particle beam, an erratic course after a beam burst would not be a definite indication of decisive damage. A certain kill against a cruise missile would be signalled by the explosion of its warhead or the plunging of the missile into the water. The missile then must be fired at until it is killed and that requires tracking capability for the beam. A missile armed with a nuclear warhead can destroy a ship even if it explodes some distance from it, and that may shorten the time available to the defence to destroy it.

The attacker can also use several countermeasures against a ship-borne charged particle beam weapon. For example, a transonic cruise missile without power and a ratio of lift to drag of 7 loses only 2% in velocity in the first 1 km of glide. In these conditions, multiple decoys could be popped out of a cruise missile, all of which would glide towards the target for the last 1 or 2 km , confusing the defence. Cruise missiles attacking a ship could use an ahead-fired smoke rocket to impair tracking by optical systems, or could fire ahead either chaff for countering radars, or even small explosive rockets for disturbing the uniformity of the air ahead, if strict uniformity is required for charged particle beam propagation.

A land-based ABM system using a charged particle beam as a kill mechanism suffers from the same basic vulnerabilities as earlier ABM schemes^{33,34} that utilised missiles as RV interceptors. The fact that a charged particle beam moves very nearly at the speed of light while missile interceptors are only hypersonic, does not remedy these vulnerabilities.

Conclusions

The current state of accelerator technology is reflected in Table 3. There is no machine which combines the high energy, high current, beam achromaticity and normalised emittance required of an accelerator suitable for a weapon system. This does not mean that accelerators with performance characteristics required for a beam weapon system could not be built. But the operational difficulties of a particle beam weapon outlined here seem insurmountable.

The cost-effectiveness of a charged particle beam system must also be compared to alternate systems, as for example, high energy lasers. Exoatmospheric uses of lasers apparently offer a more near-term operational capability, and also involve a much more available technology that one can assess and evaluate. For endoatmospheric uses, charged particle weapons are faced with competing conventional weapons such as guided projectiles, hypersonic missiles, and rapidly firing point-defence guns. Therefore, in an environment of finite resources for military research and development, it seems unwise and wasteful to develop particle beam weapons, before a detailed systems analysis of countermeasures is made. The performance characteristics of such a system do not overcome the vulnerabilities and weaknesses of earlier ABM schemes^{33,34} that used missiles as interceptors, nor do they lend themselves to the development of an effective substitute of these earlier ABM systems.

This article is the result of an informal study conducted in collaboration with G. Brandenburg (M.I.T.), R. L. Garwin (I.B.M.), and E. Knapp (Los Alamos Scientific Laboratories), whom we thank for their invaluable help.

1. Robinson, C. A. *Aviat. Week Space Technol.* **109**(15), 42 (1978); **109**(16), 42 (1978).
2. Garwin, R. L. *Bull. atom. Scient.* **34**, 24 (1978).
3. Parmentola, J. & Tsipis, K. *Scient. Am.* **240**, 54 (1979).
4. Beste, D. C. *IEEE Trans. Aerospace Electron. Syst.* **AES-14**, 446 (1978).
5. Abramowitz, M. & Stegun, I. A. *Handbook of Mathematical Functions*, 319 (Dover, 1970).
6. Alfven, H. *Phys. Rev.* **55**, 425 (1939).
7. Jackson, J. D. *Classical Electrodynamics*, Ch. 12 (Wiley, New York, 1962).
8. Dimov, G. I., Dereviankin, G. Ye & Dudnikov, V. G. *IEEE Trans. Nucl. Sci.* **NS-24**, 1545 (1977).
9. Prelec, K. (ed.), *Proc. Symp. Production and Neutralization of Negative Hydrogen Ions and Beams* (Brookhaven National Laboratory Report No. BNL 50727 UC-34a, 1977).
10. Sweetman, D. R. *Culham Laboratory Rep.* No. R112 (Abingdon, 1971).
11. Hammer, D. A. & Rostoker, N. *Phys. Fluids* **13**, 1831 (1970).
12. Davidson, R. C. *Theory of Nonneutral Plasmas* (Benjamin, New York, 1974).
13. Lawson, J. D. *J. Elect. Control* **3**, 587 (1957); **5**, 146 (1958); *J. Nucl. Energy C* **1**, 31 (1959).
14. Briggs, R. J., Clark, J., Fessenden, T., Lee, E. & Lauer, E. *Lawrence Livermore Laboratory Rep.* No. UCID-17516 (1977).
15. Lee, E. P. *Phys. Fluids* **21**, 1327 (1978).
16. Lauer, E. J., Briggs, R. J., Fessenden, T. J., Hester, R. E. & Lee, E. P. *Phys. Fluids* **21**, 1344 (1978).
17. MacDonald, A. D. *Microwave Breakdown in Gases*, 174 (Wiley, New York, 1966).
18. Link, W. T. *IEEE Trans. Nucl. Sci.* **14**, 777 (1967).
19. Rossi, B. *High Energy Particles* (Prentice-Hall, Englewood Cliffs, 1956).
20. Briggs, R. J., Hester, R. E., Lauer, E. J., Lee, E. P. & Spoerlein, R. L. *Phys. Fluids* **19**, 1007 (1976).
21. Lee, E. P. *Phys. Fluids* **19**, 60 (1976).
22. Williams, E. J. *Proc. R. Soc. A* **169**, 542 (1939).
23. Blanchard, C. H. *NBS Circ.* 527, 9 (1954).
24. Bethe, H. A. & Ashkin, J. in *Experimental Nuclear Physics* Vol. 1 (ed. Segre, E.) 168 (1953).
25. Mott, N. F. & Massey, H. S. W. *The Theory of Atomic Collisions*, 367 (Oxford Press, 1949).
26. Yonas, G. & Spence, P. Rec. 10th Symp. *Electron, Ion, and Laser Beam Technology* (ed. Marton, L.), 143, (San Francisco Press 1969).
27. Lee, E. P., Chambers, F. W., Lodestro, L. L. & Yu, S. S. *Proc. 2nd int. Conf. High-Power Electron and Ion Beam Res. and Technology*, Cornell University Vol. I, 381 (1977).
28. Charbonnier, F. M., Barbour, J. P., Brewster, J. L., Dyke, W. P. & Grundhouser, F. J. *IEEE Trans. Nucl. Sci.* **NS-14**, 789 (1967).
29. Clark, J. J. & Linke, S. *IEEE Trans. elect. Dev.* **ED-18**, 322 (1971).
30. Griem, H. R. *Plasma Spectroscopy* Ch. 1 & 2 (McGraw-Hill, New York, 1964).
31. Stier, P. M. & Barnett, C. F. *Phys. Rev.* **103**, 896 (1956).
32. Barnett, C. F. & Reynolds, H. K. *Phys. Rev.* **109**, 355 (1958).
33. Garwin, R. L. & Bethe, H. A. *Scient. Am.*, March (1968).
34. Garwin, R. L. *Internal Security* Vol. 1, No. 2 (1976).

ARTICLES

Fission track age of the KBS Tuff and associated hominid remains in northern Kenya

A. J. W. Gleadow

Department of Geology, School of Earth Sciences, University of Melbourne, Parkville, Victoria 3052, Australia

Fission track ages of pumice from the KBS Tuff have been redetermined using carefully controlled etching and counting procedures. Primary igneous zircons from five pumice samples give a mean age of 1.87 ± 0.04 Myr. The presence of detrital minerals of two different age groups has important implications for the controversy over the age of this tuff.

OVER the past 10 yr the Plio-Pleistocene sedimentary basin on the eastern shore of Lake Turkana (formerly Rudolf) in northern Kenya has been the site of an intensive study of early man and his contemporary environment. Hominid remains, other vertebrate fossils and archaeological material have all been recovered in abundance from this area and various aspects of this multidisciplinary research effort have been described recently¹⁻³.

It was realised very early that several tuffaceous horizons occurring within the sedimentary succession could be used to establish a series of radiometrically dated marker horizons⁴⁻⁷. Deposition of these tuffs in their present locations is thought to have occurred rapidly after the eruptions which produced them. Measured ages on the igneous materials can, therefore, be related directly to the stratigraphic ages of the surrounding sediments and the fossils they contain. Figure 1 gives a generalised stratigraphic column showing the relative positions of the major tuffaceous horizons in the Koobi Fora Formation of the East Turkana Basin. Also shown are the approximate positions of a few of the most important hominid and archaeological sites and their relationship to the KBS Tuff.

Results of K-Ar, ⁴⁰Ar-³⁹Ar and fission track dating studies on these tuffs have been controversial and particular problems have been associated with the KBS Tuff. Fitch and Miller⁸ originally dated feldspars from pumice in this tuff at 2.61 ± 0.26 Myr using the ⁴⁰Ar-³⁹Ar method and although further work^{9,10} produced a scatter of ages from 0.52 to 2.64 Myr this remained their preferred age. Curtis *et al.*¹¹, however, obtained mean conventional K-Ar ages of 1.82 ± 0.04 Myr and 1.60 ± 0.05 Myr respectively for tuffs mapped as the KBS in two different areas of the East Turkana Basin. Subsequently Fitch *et al.*^{12,13} revised their former estimate to 2.42 ± 0.01 Myr in apparent agreement with an average zircon fission track age of 2.44 ± 0.08 Myr reported by Hurford *et al.*¹⁴.

Note that the zircon fission track ages of Hurford *et al.* represent the first attempt at dating young zircons of this kind. These ages apparently agree closely with each other but this is

mainly due to the close communication between these authors on track identification and discrimination in these samples. Further work has not supported this consistency but initially produced a much larger scatter of results. It has become clear that this scatter was due to analytical rather than geological factors and that some of the potential difficulties had not been given adequate consideration in the earlier results.

The present study was initiated to overcome this unsatisfactory state of affairs by first establishing reliable fission track ages for zircons from the KBS Tuff. An important component in this project has been the development of improved techniques to overcome the particular problems that arise when dating zircons from geologically young rocks. The ultimate aim is to establish an independent fission track chronology for many of the tuffs in the East Turkana sedimentary basin.

Sampling and methods

None of the so-called tuffs in the East Turkana Basin represents primary pyroclastic deposits. All have been reworked by running water and redeposited as highly tuffaceous sediments. The deposits are not pure volcanic material and have been shown to contain varying amounts of detrital minerals⁷ which could seriously affect attempts at dating. Sampling was therefore restricted to the occasional blocks of pumice (up to ~25 cm across) which are found in some of the tuffs as these represent relatively uncontaminated igneous material. Pumice samples were collected from the KBS Tuff in Fossil Collection Areas 105, 112 and 131 which are shown in Fig. 2. The detailed petrography and mineralogy of some of these samples has been described elsewhere¹⁵. The pumice from Areas 105 and 131 was found to contain small accessory zircon crystals (up to ~450 × 150 μm but mostly much smaller) suitable for fission track dating. No primary igneous zircons were found in pumice from Area 112.

One of the zircon separates originally dated by Hurford *et al.* (FMA517) was reanalysed in this study. These zircons were extracted from a sample of several pumice cobbles by dissolving

the crushed pumice in a mixture of HF and H₂SO₄ and recovering the crystals from the insoluble residue¹⁴. Zircons were extracted from two large single pumice blocks, FMA559 and FMA560, in essentially the same way. A different separation procedure was used for the two remaining samples (7722-numbers) both of which were composite samples of a large number of pumice cobbles. After an initial coarse crushing to <2 mm a panning procedure was used to separate crystals from the pumice. Zircons were then extracted from the phenocryst concentrate using conventional heavy liquid and magnetic techniques and an additional fraction was obtained by acid dissolution of the glass. Two other minerals suitable for fission track dating, apatite and sphene, were also recovered from the pumices but these were found to be almost entirely of detrital origin.

Fission track dating methods for zircons were based on those described by Gleadow *et al.*¹⁶ and the external detector approach for measuring induced tracks was used. Because of the very young geological age of these zircons, particular difficulties arise which are not generally encountered with older materials. As most previous experience with dating zircons has been with much older rocks than those studied here, these difficulties and their potential effect on the resulting ages have not been clearly understood. Some of the factors involved are as follows:

- (1) Very small numbers of fossil fission tracks are found in the zircons. This is exaggerated in the present samples by a much lower than average uranium concentration and often small grain size. The lack of tracks makes it difficult to assess the correct etching time for these zircons.
- (2) The zircons have a very low track etching rate due to the low level of α -radiation damage¹⁶ making it difficult to reveal all the tracks by etching. Also, the zircon grains tend to fall out of their FEP Teflon mounts during the long etching times required.
- (3) Etching of tracks lying in different orientations is highly anisotropic giving rise to etch pits of differing appearance and making identification more difficult.
- (4) Tiny acicular inclusions are present in the zircons from the KBS Tuff which are often of very similar dimensions to etched tracks and can easily be mistaken for tracks.

The anisotropic etching of fission tracks in different crystallographic orientations in zircons with low radiation damage was examined in more detail by progressively etching collimated ²⁵²Cf tracks in a large zircon crystal from the Mud Tank Carbonate in Central Australia. This zircon has a spontaneous track density of about $3 \times 10^6 \text{ cm}^{-2}$ and a correspondingly high level of radiation damage. One part of the zircon crystal was annealed at 850 °C for 1 h to simulate the absence of significant α -damage in the young zircons and to remove all preexisting fission tracks. The ²⁵²Cf tracks were then implanted on a polished (110) cleavage face in two orientations, parallel and perpendicular to the *c*-axis, both dipping at 30° to the surface. Figure 3 shows how one track etching parameter, the track density, varies with

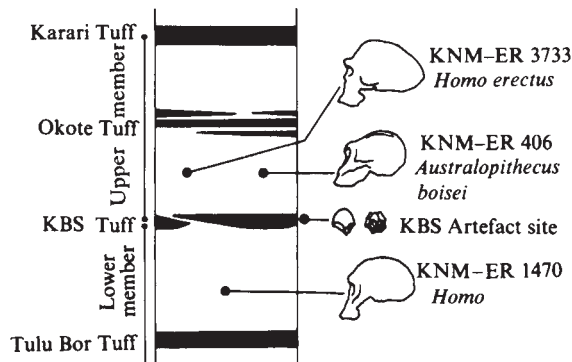


Fig. 1 Generalised stratigraphic column showing tuffs occurring in the Koobi Fora Formation to the east of Lake Turkana in northern Kenya. The approximate positions of some of the most important hominid and archaeological sites are also shown.

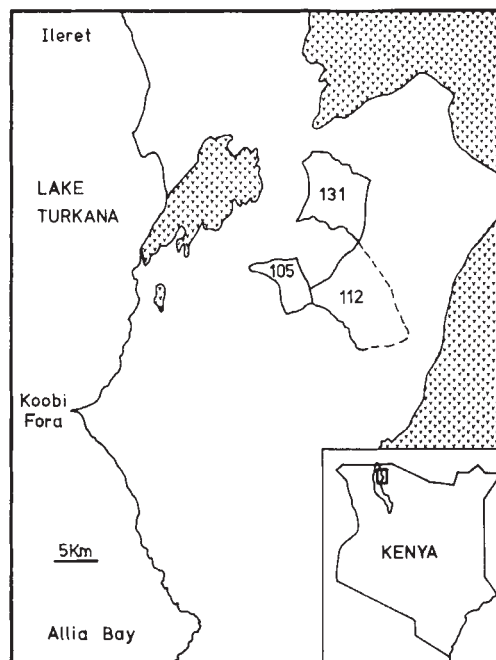


Fig. 2 Locality map of the East Turkana Basin showing the numbered fossil collection areas where the fission track dating samples were collected. Shaded areas indicate the older volcanic basement.

etching time for the two sets of ²⁵²Cf tracks compared with that for spontaneous tracks in an identical but unannealed sample of the zircon. The very different etching rates for the differently oriented tracks is obvious. The etched track shapes are also very different for the two orientations. Tracks perpendicular to the *c*-axis etch more readily and may appear quite satisfactory for counting before many of the tracks parallel to the *c*-axis are yet visible. With continued etching the *c*-parallel tracks become more like the acicular inclusions in appearance and great care is needed to discriminate between the two. These experiments enabled criteria to be set up for the reliable identification of etched tracks in all orientations and for their discrimination from spurious track-like features.

A small test mount was then made up of each of the zircons to be dated and this was irradiated with a high density of ²⁵²Cf tracks. This test mount was then etched alongside the main dating mount for each sample and was inspected periodically to judge the etching time required. Etching was continued for sufficient time (up to 100 h in KOH:NaOH eutectic at 210 °C) to ensure that the tracks in the slow-etching orientations were adequately revealed. The appearance of the etched ²⁵²Cf tracks in the auxiliary mount was then used as a guide to the expected appearance of fossil tracks in equivalent crystallographic orientations. This enabled a much more rigorous discrimination to be made between tracks and spurious features. A test of these discrimination procedures was made by 'dating' a sample of pumice zircon (FMA560) which had been annealed at 850 °C for 1 h to remove all spontaneous tracks.

Another significant departure from standard practice with the external detector method was to count every grain in the mount or in a predetermined area of the mount which was acceptable on the usual surface-orientation and etching criteria¹⁷. This avoided the selection of a small number of grains (typically only about five) for counting which is usually made with older (high track density) zircons. While this selection is inconsequential with high spontaneous track densities, it introduces a serious risk of bias where there are only a few tracks per grain or even no tracks in some cases. In addition all fission track counting of all samples was completed before any ages were calculated to guard against the possibility of earlier results influencing the identification of tracks in later determinations. After ~12 months the entire counting procedure, including the choice of

grains to be counted, was repeated for most of the samples as independently as possible from the first count. In this way duplicate, though not totally independent, results were obtained.

Neutron dosimetry procedures and the techniques used for apatite and sphene have been described in detail elsewhere¹⁸. Flux gradients during irradiation were monitored by using two dosimeter glasses. Approximately 1,200–1,600 tracks were counted over each dosimeter. Constants used in the age calculations were: $\lambda_F = 6.9 \times 10^{-17} \text{ yr}^{-1}$, $\lambda_D = 1.551 \times 10^{-10} \text{ yr}^{-1}$, $I = 7.253 \times 10^{-3}$, $^{235}\sigma = 5.802 \times 10^{-22} \text{ cm}^2$. Errors were calculated according to the method of Naeser *et al.*¹⁹ and McGee and Johnson²⁰ and are expressed as 1 standard deviation (σ) throughout.

KBS Tuff primary zircons

Analytical results are shown in Table 1 for primary igneous zircons from two individual and three composite pumice samples from the KBS Tuff. Results are also shown for two standard zircons of independently known age which were included in the same neutron irradiations as some of the KBS zircons. It can be seen in Table 1 that the fission track ages for these zircons from the Fish Canyon Tuff, Colorado (72N8) and the high level Mt Dromedary Intrusion in New South Wales (7822-9) are in excellent agreement with their K–Ar ages calculated using the new constants recommended by Steiger and Jäger²¹.

As expected no features which would definitely be counted as spontaneous tracks were found in the annealed sample of

FMA560 zircon. After annealing this zircon was subject to identical etching, irradiation and counting procedures as the other KBS Tuff zircons. Applying the same criteria as used for the other samples to discriminate tracks from spurious etch-pits, only four features were seen over 15 grains which might possibly have been confused with tracks and none of these were convincing. These would give a maximum contribution of only $0.11 \pm 0.05 \text{ Myr}$ to the age if all had been mis-identified as tracks. As even this background level is considered unlikely this experiment gives greater confidence that a reliable discrimination has been made for the dated KBS zircons.

For two of the KBS pumices (FMA559 and FMA560) sufficient zircon was obtained for more than one mount to be prepared and these were all etched and irradiated separately. Thus three independent age measurements were obtained for FMA559 and two for FMA560 giving eight measurements for the five pumices. Six of these were counted twice with very close agreement between the duplicate measurements in all but one case. For the second irradiation of FMA560 the duplicate results, which were counted over two quite separate sets of grains, are significantly different from each other at the 95% confidence level. The lower result, $1.54 \pm 0.10 \text{ Myr}$, seems anomalous. Excluding this result the zircon ages for the two individual pumice blocks from Area 131 are indistinguishable from each other and give a mean age and standard deviation of $1.85 \pm 0.06 \text{ Myr}$. This mean age is also indistinguishable from the ages of $1.85 \pm 0.06 \text{ Myr}$ and $1.90 \pm 0.08 \text{ Myr}$ obtained for the composite pumice samples from Areas 131 and 105 respec-

Table 1 Fission track ages of zircons from the KBS Tuff pumices and age standards

Rock no.	Area	No. of grains	<i>R</i>	$\rho_s \times 10^4$ (cm ⁻²)	$\rho_i \times 10^6$ (cm ⁻²)	$n \times 10^{15}$ (n cm ⁻²)	U (p.p.m.)	Calculated age (Myr ± 1σ)	Average age (Myr ± 1σ)
Individual pumices									
FMA559	131	14	0.666	8.78 (68)	3.64 (1411)	1.31	101	1.93 ± 0.21	1.79 ± 0.13
FMA559	131	19	0.944	9.22 (82)	4.30 (1910)	1.31	119	1.71 ± 0.16	
FMA559	131	26	0.847	6.77 (115)	3.00 (2583)	1.37	80	1.86 ± 0.15	1.86 ± 0.15
FMA559	131	22	0.928	7.59 (193)	5.19 (6597)	2.05	92	1.83 ± 0.12	1.83 ± 0.12
FMA560	131	12	0.934	13.1 (156)	5.36 (3196)	1.31	149	1.95 ± 0.14	1.91 ± 0.10
FMA560	131	20	0.937	11.4 (143)	4.86 (3046)	1.31	135	1.87 ± 0.14	
FMA560	131	15	0.892	8.58 (91)	5.73 (3039)	2.00	104	1.83 ± 0.17	1.62 ± 0.09
FMA560	131	24	0.973	8.62 (177)	6.84 (7019)	2.00	125	1.54 ± 0.10	
Composite pumices									
FMA517	131	17	0.822	6.26 (137)	2.96 (3241)	1.42	76	1.83 ± 0.14	1.83 ± 0.10
FMA517	131	20	0.850	6.29 (139)	3.00 (3313)	1.42	77	1.82 ± 0.13	
7722-107	131	15	0.943	13.50 (199)	8.40 (6199)	1.97	155	1.93 ± 0.12	1.88 ± 0.09
7722-107	131	15	0.936	12.5 (173)	8.24 (5688)	1.97	152	1.82 ± 0.12	
7722-108	105	15	0.794	9.48 (121)	4.52 (2886)	1.44	114	1.84 ± 0.15	1.90 ± 0.08
7722-108	105	27	0.935	9.87 (277)	4.49 (6301)	1.44	114	1.93 ± 0.10	
Annealed zircon									
FMA560	131	15	0.184	<0.424 (?4)	4.81 (2269)	1.98	89	<0.11 ± 0.05	
Age standards									
72N8	—	6	0.912	623 (602)	28.3 (1367)	2.08	495	27.9 ± 0.9	K–Ar age 27.8 ± 0.4*
7822-9	—	5	0.807	2190 (1475)	23.5 (1587)	1.75	489	98.7 ± 2.6	97.9 ± 0.7†

R = correlation coefficient for spontaneous and induced track counts; ρ_s = spontaneous track density; ρ_i = induced track density; n = neutron dose; U = uranium concentration. Figures given in parentheses under each track density show number of tracks counted.

* Mean K–Ar age on biotite, hornblende, sanidine and plagioclase²⁶.

† Biotite K–Ar age²⁷.

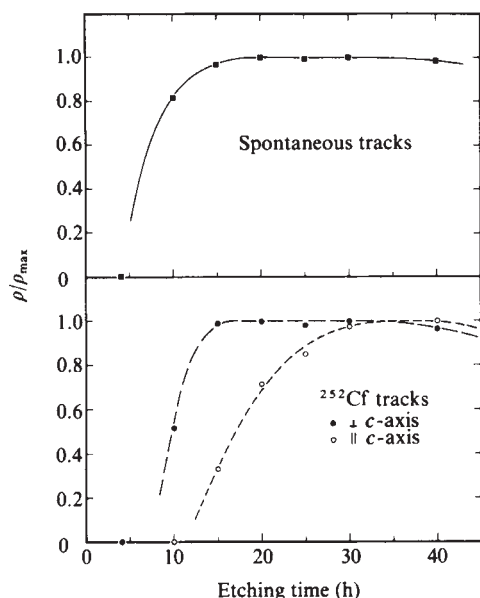


Fig. 3 Variation in observed track density (ρ) with etching time for spontaneous fission tracks and oriented ^{252}Cf tracks in an annealed sample of the same zircon. The track densities are expressed as a function of ρ_{max} , the maximum track density revealed.

tively. The close concordance of ages for the composite samples and those for individual pumice blocks strongly suggests that zircon-bearing pumice of only one age is present in the KBS Tuff.

A possible explanation for the anomalous FMA560 result of 1.54 ± 0.10 Myr may lie in the extreme range of uranium concentrations found in this sample. The 24 grains counted for this determination ranged from 54 to 516 p.p.m. uranium which is greater than for any other sample. This wide range in uranium concentrations is reflected in the high correlation coefficient observed in this case. It has been shown¹⁶ that the track etching rate in zircon varies with the total radiation damage so that

low-U grains etch more slowly than high-U grains for the same age. Thus it is possible that this determination included some low-U grains which were under-etched thereby depressing the combined age. The combined age of just the high-U grains alone (those with greater than the average of 125 p.p.m.) is 1.85 ± 0.17 Myr which supports this explanation.

The locality sampled in Area 105 can be traced directly to the type locality for the KBS Tuff but outcrop of this tuff is discontinuous between Areas 105 and 131. Nevertheless, the correlation is considered quite firm on the basis of stratigraphy⁷ and the compositions of pumice glasses²² and phenocrysts¹⁵. The dating results reported here also confirm the correlation between the two areas and, excluding the anomalous FMA560 value, give a mean age of 1.87 ± 0.04 Myr which is regarded as the best estimate of the fission track age of the KBS Tuff. Inclusion of the anomalous age would lower this mean to 1.83 ± 0.04 Myr which is not significantly different. The preferred mean fission track age of 1.87 ± 0.04 Myr is in close agreement with the mean K-Ar age of 1.89 ± 0.01 Myr reported by McDougall *et al.*²³ and that of 1.83 ± 0.06 Myr calculated from the data of Drake *et al.*²⁴ (all 1σ errors).

Detrital contamination

Some detrital contamination was present in all zircon concentrates dated, presumably from the small amounts of tuffaceous matrix which had worked deep into vesicles and fractures in the pumice. One of the major advantages of the external detector method of fission track dating, however, is that it permits the determination of ages on individual mineral grains, even at very small grain size ($<100 \mu\text{m}$). Thus it is possible to discriminate between the young primary igneous zircons and the old detrital zircons.

Two distinct groups of detrital zircons have been identified in the KBS Tuff and apart from these no other detrital grains have been found. The first consists of very rare zircon crystals which are euhedral and give single crystal ages between 24 and 27 Myr as shown in Table 2. Zircons of this age group were also found in mineral concentrates from pumices in the Tulu Bor Tuff, the Karari and its correlative, the Chari Tuff. Two apatite crystals of closely similar fission track age were found in pumice from the KBS Tuff in Area 105 (Table 2). These inherited zircons are thought to have been derived from a volcanic source probably to

Table 2 Fission track ages of detrital minerals occurring in pumices from the KBS Tuff

Rock no. 7722-	Area	Mineral	No. of grains	<i>R</i>	$\rho_s \times 10^6$ (cm^{-2})	$\rho_i \times 10^6$ (cm^{-2})	$n \times 10^{15}$ ($n \text{ cm}^{-2}$)	U (p.p.m.)	Age (Myr $\pm 1\sigma$)
'Volcanic' detritals									
108	105	Apatite	1	—	0.210 (18)	2.15 (92)	4.98	17	29.7 ± 8
108	105	Apatite	1	—	0.113 (4)	1.35 (24)	4.98	11	25.4 ± 14
108	105	Zircon	1	—	0.527 (34)	1.95 (63)	1.44	49	23.7 ± 5
108	105	Zircon	1	—	0.465 (24)	1.51 (39)	1.44	38	27.0 ± 7
108	105	Zircon	1	—	1.33 (180)	4.70 (318)	1.44	119	24.8 ± 2
'Basement' detritals									
108	105	Apatite	10	0.965	0.763 (504)	2.78 (918)	4.98	22	83 ± 3
108	105	Zircon	3	0.823	11.3 (399)	1.73 (918)	1.44	44	549 ± 32
108	105	Sphene	5	0.993	7.42 (813)	5.58 (611)	6.47	34	504 ± 10
109B	112	Zircon	6	0.887	19.5 (1078)	4.14 (407)	1.94	78	534 ± 17
109C	112	Zircon	3	0.999	29.6 (598)	3.07 (297)	1.92	58	456 ± 12
109C	112	Sphene	5	0.969	8.04 (830)	5.90 (609)	6.04	39	483 ± 11
107	131	Zircon	3	0.917	10.5 (711)	2.63 (237)	1.97	49	462 ± 22

Table 3 Fission track dating results for rocks from the Mozambique Belt basement in Kenya

Rock no.	Rock type	Mineral	No. of grains	<i>R</i>	$\rho_s \times 10^6$ (cm ⁻²)	$\rho_i \times 10^6$ (cm ⁻²)	$n \times 10^{15}$ (n cm ⁻²)	U (p.p.m.)	Age (Myr $\pm 1\sigma$)
100	Gneiss	Sphene	5	0.940	10.0 (2901)	6.77 (982)	5.52	49	479 $\pm$ 12
101	Gneiss	Apatite	6	0.932	0.243 (365)	1.85 (1386)	8.75	8	70 $\pm$ 2
102	Gneiss	Apatite	6	0.828	0.253 (334)	1.57 (1035)	8.69	7	85 $\pm$ 3
103	Gneiss	Zircon	6	0.994	8.42 (1324)	1.21 (95)	1.26	35	514 $\pm$ 40
103	Gneiss	Apatite	6	0.965	0.338 (446)	2.20 (1456)	8.63	10	80 $\pm$ 2
104	Gneiss	Apatite	7	0.937	0.328 (533)	2.75 (2237)	8.56	13	62 $\pm$ 2
136	Mag-Ap rock	Apatite	5	0.693	0.631 (1133)	3.27 (1470)	8.44	15	99 $\pm$ 3
141	Granite gneiss	Sphene	5	0.950	20.5 (1621)	11.2 (1177)	5.07	88	542 $\pm$ 12
141	Granite gneiss	Apatite	8	0.945	1.04 (484)	4.09 (952)	8.32	19	127 $\pm$ 3

the east or north-east. No detrital ages younger than 24 Myr were found in zircon or apatite concentrates from the KBS Tuff. Miocene volcanics around the southeastern margin of the East Turkana Basin have given ⁴⁰Ar–³⁹Ar ages of 7.5 and 11.8 Myr (ref. 10) but a collection of these rocks proved to be almost totally lacking in zircon so that these are not a potential source of detrital contamination.

The other group of detrital zircons are much more abundant, often making up ~10–20% of these zircon concentrates. These are readily identified by their rounded, water-worn shapes, frequently darker colour and generally high spontaneous track densities. Two other minerals, sphene and apatite, with similar characteristics to these zircons are also found as detrital minerals in the pumice samples. Many of the detrital zircons are at least partly metamict and therefore unsuitable for fission track dating but satisfactory ages were obtained for some of the grains with low uranium concentrations. Table 2 shows fission track ages for all these detrital minerals, each age representing the combined counting results from several similar grains. Zircon and sphene ages from this group all fall in the range 500 $\pm$ 50 Myr which is typical of radiometric ages for minerals from the crystalline rocks of the Mozambique Belt of East Africa²⁵. This is one of the 'Pan-African' mobile belts of late Precambrian–early Palaeozoic age and form the basement complex of the northern Kenya region.

Detrital apatites from the KBS Tuff in Area 105 (7722–108) give a combined age of 83 $\pm$ 5 Myr which is not obviously related to either the 'Pan-African' detrital ages or the younger 'volcanic' detrital ages. However, apatites frequently give much younger fission track ages than sphenes or zircons from the same rocks due to the much lower thermal stability of tracks in apatite. Table 3 shows ages for apatite, sphene and zircon separated from basement gneisses of the Mozambique Belt from various parts of Kenya as shown in Fig. 4. It can be seen that sphene and zircon both show typical 'Pan-African' ages (500 $\pm$ 50 Myr) but apatites give much younger ages ranging from 62 to 127 Myr with a weighted mean of 82 Myr. This pattern is almost identical to that found in the East Turkana detrital minerals. Thus it seems that the old detrital component found in the pumice samples has probably been derived from basement rocks of the Mozambique Belt.

Significance of the results

Two conflicting hypotheses have been offered to explain the different K–Ar and ⁴⁰Ar–³⁹Ar feldspar ages obtained for the KBS Tuff. Fitch *et al.*¹² consider the younger ages (~1.8 Myr) to be the result of a partial or total over-printing event. Curtis *et al.*¹¹, on the other hand, regard the older ages (~2.4 Myr) to be due to contamination of the feldspar separates by older material. It has also been pointed out that individual pumice blocks within

each of the reworked tuffs could conceivably be from more than one source and possibly of more than one age. Thus different ages might be obtained for a variety of analytical and geological reasons from samples collected at the one locality. K–Ar evidence of Curtis *et al.*¹¹ suggesting that tuffs mapped as the KBS in Areas 105 and 131 were of slightly different age, has now been eliminated with the discovery of a systematic error in the lower (1.6 Myr) ages²⁴.

The fission track evidence presented here has an important bearing on this discussion. There is very obvious agreement between the KBS Tuff zircon ages in Table 1 and the K–Ar ages of 1.8–1.9 Myr obtained by Curtis *et al.*¹¹, Drake *et al.*²⁴ and by McDougall *et al.*²³ which suggests that these ages date a real geological event. Also no fission track evidence was found for pumice of more than one age in the KBS Tuff from Area 131. Similarly there is no evidence that pumice from the KBS Tuff in Area 105, the type area for this tuff, has a different age from its

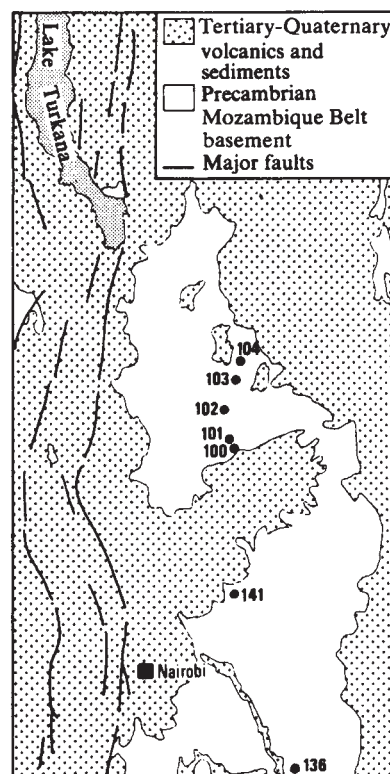


Fig. 4 Simplified geological map of part of Kenya showing localities of Precambrian basement rocks collected for fission track dating. Sample numbers have the prefix 7722–.

correlative in Area 131. All these observations argue against a purely geological explanation for the discrepancies in the published ages.

Table 2 clearly demonstrates the presence of detrital zircon, sphene and apatite grains in mineral concentrates separated from the pumice samples, even after very careful sample preparation. It therefore seems highly likely that feldspars separated for K–Ar dating could also contain traces of much older basement feldspar. This supports the contention that older K–Ar and ^{40}Ar – ^{39}Ar ages are the result of contamination. As discussed above the fission track ages of Hurford *et al.*¹⁴ are thought to be too old for purely analytical reasons, in particular the mis-identification of a finite number of acicular inclusions or dislocations as tracks and possibly a biased choice of grains for counting.

The presence of old detrital fission track ages also argues very strongly against the disturbance of any of the dating systems by postulated overprinting events. The pattern of ages found for detrital zircon, sphene and apatite closely matches that found in rocks of the Mozambique Belt basement which shows that the fission track ages have not been disturbed during erosion, deposition or any subsequent event. Considering the relative sensitivities of the different mineral dating systems to overprinting, it is inconceivable that total overprinting of glass and feldspar K–Ar ages at 1.8 Myr, as suggested by Fitch *et al.*¹², could have occurred without drastic effects on the detrital fission track ages, particularly those of the apatites. Thus the preservation of the old basement ages shows the hypothesis of post-depositional overprinting events to be unreasonable.

Received 25 October; accepted 17 December 1979.

1. Coppins, Y., Howell, F. C., Isaac, G. L. & Leakey, R. E. F. (eds) *Earliest Man and Environments in the Lake Rudolf Basin* (University of Chicago Press, 1976).
2. Bishop, W. W. (ed.) *Geological Background to Fossil Man* (Scottish Academic Press, Edinburgh, 1978).
3. Leakey, M. G. & Leakey, R. E. F. (eds) *Koobi Fora Research Project Vol. 1* (Clarendon Press, Oxford, 1978).
4. Behrensmeyer, A. K. *Nature* **226**, 225 (1970).
5. Bowen, B. E. & Vondra, C. F. *Nature* **242**, 391 (1973).
6. Findlater, I. C. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppins, Y. *et al.*) 94 (University of Chicago Press, 1976).
7. Findlater, I. C. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 415 (Scottish Academic Press, Edinburgh, 1978).
8. Fitch, F. J. & Miller, J. A. *Nature* **226**, 226 (1970).
9. Fitch, F. J., Findlater, I. C., Watkins, R. T. & Miller, J. A. *Nature* **251**, 213 (1974).
10. Fitch, F. J. & Miller, J. A. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppins, Y. *et al.*) 123 (University of Chicago Press, 1976).
11. Curtis, G. H., Drake, R., Cerling, T. & Hampel, J. *Nature* **258**, 395 (1975).
12. Fitch, F. J., Hooker, P. J. & Miller, J. A. *Nature* **263**, 740 (1976).
13. Fitch, F. J., Hooker, P. J. & Miller, J. A. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 444 (Scottish Academic Press, Edinburgh, 1978).

Conclusions

Redetermination of the fission track age of primary igneous zircons from pumice contained in the KBS Tuff gives a mean age of 1.87 ± 0.04 Myr. This is regarded as representing the time of eruption of the pumice which in turn is taken to be a close approximation to the time of the deposition of this tuff in the East Turkana Basin. Individual pumice blocks from the KBS Tuff in Area 131 gave ages indistinguishable from each other and from pumice in Area 105.

The external detector method of fission track dating is particularly advantageous in detecting detrital contamination and discriminating against its effects on resulting ages. Two detrital components were identified, one showing essentially concordant zircon and apatite ages and probably from a volcanic source of dominantly early Miocene age. The other has a strongly discordant pattern of apatite ages compared with sphene and zircon which matches that observed for the basement rocks of the Mozambique Belt. The preservation of these detrital mineral ages fully justifies contamination as a plausible explanation for the older K–Ar ages and is strong evidence against the existence of previously postulated overprinting events.

This work was supported by a Queen Elizabeth II Fellowship and in part by a grant from the Australian Institute of Nuclear Science and Engineering. I thank R. E. F. Leakey for his encouragement and for supporting the fieldwork, also A. J. Hurford for some zircon separates and T. E. Cerling and I. C. Findlater for valuable guidance and discussion.

14. Hurford, A. J., Gleadow, A. J. W. & Naeser, C. W. *Nature* **263**, 738 (1976).
15. Ferguson, A. K. & Gleadow, A. J. W. in *Geodynamic Evolution of the Afro-Arabian Rift System* (in the press).
16. Gleadow, A. J. W., Hurford, A. J. & Quaipe, R. D. *Earth planet. Sci. Lett.* **33**, 273 (1976).
17. Gleadow, A. J. W. *Nucl. Track Detect.* **2**, 105 (1978).
18. Gleadow, A. J. W. & Lovering, J. F. *J. geol. Soc. Aust.* **25**, 323 (1978).
19. Naeser, C. W., Johnson, N. M. & McGee, V. E. *U.S. Geol. Surv. Open-File Rep.* **78-701**, 303 (1978).
20. McGee, V. E. & Johnson, N. M. *Math. Geol.* **11**, 255 (1979).
21. Steiger, R. H. & Jager, E. *Earth planet. Sci. Lett.* **36**, 359 (1977).
22. Cerling, T. E., Brown, F. H., Cerling, T. E., Curtis, G. H. & Drake, R. E. *Nature* **279**, 118 (1979).
23. McDougall, I., Maier, R., Sutherland-Hawkes, P. & Gleadow, A. J. W. *Nature* **284**, 230–234 (1980).
24. Drake, R. E., Curtis, G. H., Cerling, T. E., Cerling, B. W. & Hampel, J. (in preparation).
25. Cahen, L. & Snelling, N. J. *The Geochronology of Equatorial Africa* (North-Holland, Amsterdam).
26. Steven, T. A., Mehnert, H. H. & Obradovich, J. D. *U.S. Geol. Surv. Prof. Pap.* **575-D**, 47 (1967).
27. McDougall, I. & Roksandic, Z. *J. geol. Soc. Aust.* **21**, 81 (1974).

K–Ar age estimate for the KBS Tuff, East Turkana, Kenya

Ian McDougall, Robyn Maier & P. Sutherland-Hawkes

Research School of Earth Sciences, Australian National University, Canberra, ACT 2600, Australia

A. J. W. Gleadow

Department of Geology, School of Earth Sciences, University of Melbourne, Parkville, Victoria 3052, Australia

Stone tools and numerous vertebrate fossils including hominids, have been found in close stratigraphic proximity to the KBS Tuff, whose age has been the subject of much debate. Concordant K–Ar ages, averaging 1.89 ± 0.01 Myr, are reported on anorthoclase phenocrysts from 13 pumice clasts collected from within the KBS Tuff or its correlatives. We believe that this age is the best estimate currently available for the time of formation of this important marker horizon within the East Turkana Basin.

PLIOCENE to Pleistocene sedimentary rocks which crop out adjacent to the eastern shores of Lake Turkana, northern Kenya, have yielded a wealth of vertebrate fossils, including hominids, and archaeological material^{1–3}. Rhyolitic tuffaceous beds within the sequence have facilitated mapping of the area^{1–6}, and these horizons contain samples amenable to

radiometric dating. The KBS Tuff^{4–6}, within the Koobi Fora Formation, is one of these marker beds. As important hominid and artefact finds have been made within or adjacent to this stratigraphic level^{1–3,7,8}, it is of considerable importance to obtain precise and accurate ages for the KBS Tuff. Conventional K–Ar, ^{40}Ar / ^{39}Ar and fission track dating of pumice clasts within

this tuff have yielded a distressingly large range of ages⁹⁻¹⁴. In an attempt to resolve the controversy surrounding the age of the KBS Tuff we present here the results of a conventional K-Ar dating study of anorthoclase phenocrysts separated from pumice clasts found within the tuff.

KBS Tuff

Lake Turkana lies in a basin within the East African Rift System. During the late Pliocene and the Pleistocene the northeastern part of the Lake Turkana Basin was occupied by a large embayment in which sediments accumulated^{15,16} (Fig. 1). These sediments, about 325 m thick in all, are essentially flatlying, except locally, and represent a range of facies ranging from lacustrine through deltaic to fluvial.

The KBS Tuff is the topmost unit of the lower member of the Koobi Fora Formation^{4-6,15-18}, and is a medium to coarse grained vitric tuff of variable thickness, commonly about 1 m thick. Total thickness of overlying sediment does not exceed 125 m (ref. 18). In common with most tuffaceous horizons within the East Turkana sequence, the KBS Tuff is not a primary air fall tuff, but has been transported by, and deposited from, water¹⁸. Varying proportions of detrital material, derived from Miocene volcanic rocks and early Palaeozoic and Precambrian basement terrains marginal to the Lake Turkana Basin, are found within the tuffs.

Pumice clasts, up to 25 cm across, commonly water-rounded, occur locally within the KBS Tuff, and are regarded as products of the same volcanic eruptions that comprise the bulk of the tuff^{15,17,18}. The time between eruption of the primary volcanic material, including the pumice clasts and deposition within the East Turkana Basin, is considered to be very short—virtually instantaneous in terms of dating the deposits^{17,18}. Because the pumice clasts are less likely to be contaminated by detrital material compared with the enclosing tuff, most of the earlier age measurements were made on such clasts, and this is the procedure followed in this study. The clasts consist mainly of vesicular glass with variable proportions, normally less than 10% by volume, of anorthoclase phenocrysts up to 5 mm in size, and less common pyroxene. The glass is fresh to somewhat altered, and in some cases is partly replaced by carbonate. The phenocrysts are fresh and unaltered, the feldspar normally being limpid and euhedral. In the present study only the feldspar (anorthoclase) phenocrysts have been used for the dating.

Previous dating studies

Fitch and Miller⁹ dated a feldspar phenocryst concentrate, prepared from a pumice clast in the KBS Tuff, at 2.61 ± 0.26 Myr, subsequently¹² revising the age to 2.42 ± 0.01 Myr, based on results from four steps of a $^{40}\text{Ar}/^{39}\text{Ar}$ incremental heating experiment. They¹⁰ also reported $^{40}\text{Ar}/^{39}\text{Ar}$ total fusion ages on 10 feldspar concentrates from pumice clasts in the tuff, with results ranging from 0.52 ± 0.33 to 2.64 ± 0.29 Myr. In addition, conventional K-Ar ages of 8.43 ± 0.51 and 17.5 ± 0.9 Myr on feldspar from another clast (FM7054) were given¹⁰, but these old ages were thought to be caused by contamination from detrital material. Age spectra determined by the $^{40}\text{Ar}/^{39}\text{Ar}$ dating method were reported^{10,12} for feldspar concentrates from eight clasts from the KBS Tuff. Although the age spectra were stated to be complex, these authors interpreted apparent ages of 2.4–2.5 Myr for two samples (FMA274, FMA517) as crystallisation ages, and they suggested that overprinting could be recognised in most spectra at ~1.8–2.0 Myr, and in two cases at ~1.0 Myr. Support for this age of crystallisation of the pumices found within the KBS Tuff was given by Hurford *et al.*¹³, who reported fission track ages on zircon with a mean value of 2.44 ± 0.08 Myr.

Disregarding four conventional K-Ar ages on feldspar from pumice clasts in the KBS Tuff in the range 2.01–6.9 Myr, thought to be caused by detrital contamination, Curtis *et al.*¹¹ obtained concordant K-Ar ages on feldspar and glass from pumice clasts found in this horizon with mean values of 1.82 ± 0.04 Myr and 1.60 ± 0.05 Myr in two different areas, respec-

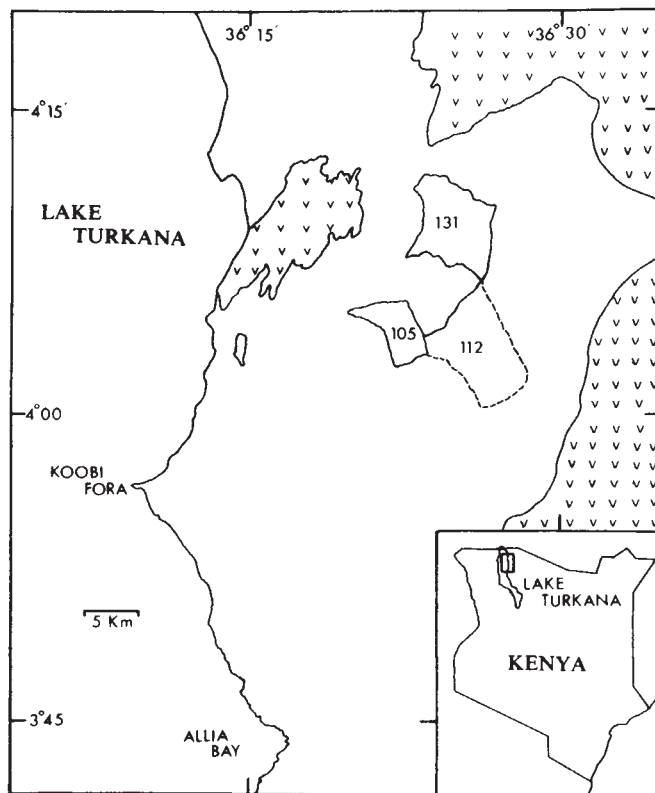


Fig. 1 Map of the Koobi Fora Region, adjacent to the northeastern shores of Lake Turkana, Kenya, showing location of the areas from which the samples for dating were obtained within the Pliocene-Pleistocene stratigraphic sequence (blank area). Volcanics forming the local basement are indicated by pattern. After Leakey and Leakey³.

tively. Subsequently, Drake *et al.*¹⁴ reported an error in potassium determinations on the samples previously dated by them¹¹ that yielded the 1.6 Myr ages. They listed¹⁴ new and revised K-Ar ages for 16 feldspar and six volcanic glasses from 13 pumice clasts in the KBS Tuff, ranging from 1.73 to 1.92 Myr, forming a reasonably concordant set with a mean of 1.8 ± 0.1 Myr.

Methods

Feldspar (anorthoclase) phenocrysts were separated from single pumice clasts found within the KBS Tuff and its correlatives in areas 105, 112 and 131, East Turkana (Fig. 1). In most cases the surface of the clasts was removed by sawing before crushing. Following crushing, anorthoclase was concentrated by a panning procedure, and normally the +1 mm fraction was retained and further crushed to $-360 + 170 \mu\text{m}$ or $-360 + 100 \mu\text{m}$. Tetra-bromoethane, adjusted to appropriate density, was used to purify the anorthoclase. The final step involved treatment of the concentrate in dilute HF (7%) ultrasonically for 5 min. Grain counting showed that the separates had a grain purity of better than 99.5%, and that glass remaining on crystal faces or as small inclusions rarely exceeded 1%.

Potassium was determined by flame photometry using Li internal standard and Na buffer on an IL 443 machine, following procedures similar to those previously described¹⁹. Argon extractions were made on aliquants of 0.9–2.9 g in a high vacuum line by radiofrequency heating. Temperatures of ~1,500°C were maintained for 30 min to ensure complete outgassing of argon from the melt. Previously it had been found²⁰ that temperatures substantially greater than that needed for fusion were required to remove all argon from K-feldspars. A tracer of ^{38}Ar , prepared from a gas pipette system, was added during the fusion. Subsequent to purification of the argon, its isotopic composition was measured in a AEIMS10 or a Micromass 12 mass spectrometer, both

operated in the static mode. Peak hopping is used on both machines and results are taken digitally on-line with a Hewlett-Packard 21MX computer, in which all data processing is done to produce an age within minutes of completion of the isotopic analysis.

Errors are quoted at the level of one standard deviation as previously described^{21,22}. The average age given in those cases where the sample was measured in duplicate has a minimum assigned error of 1%, because the uncertainties in the measurement of potassium and calibration of the ³⁸Ar tracer are of this order.

All ages are calculated using the revised ⁴⁰K decay constants and isotopic abundances, recently recommended²³. These constants yield ages 2.67% greater than those reported by most previous workers, but Drake *et al.*¹⁴ have also used the new constants.

Results and discussion

Petrographically the K-feldspar phenocrysts in the pumice clasts are homogeneous with no evidence of unmixing. Extremely fine polysynthetic twinning is observed in some crystals. X-ray diffraction measurements, using the three peak method of Wright²⁴, carried out by W. D. Birch (personal communication), indicate that the feldspar is highly disordered and structurally close to the sanidine-high albite series, exhibiting no evidence for significant unmixing or strain resulting from exsolution. The feldspar has a composition in the range Or₃₅–Or₃₉ (weight %) from X-ray diffraction analysis, comparing well with the composition of Or₃₅–Or₃₇ derived from the potassium analyses (Table 1). Following the nomenclature of Wright and Stewart²⁵ these feldspars are classified as anorthoclase.

Results of the K–Ar dating are given in Table 1. Duplicate potassium measurements agree to better than 1%, and the range of values is limited to less than 4% relative. In most cases sufficient anorthoclase was available to enable argon to be measured in duplicate; generally these agree to better than 2%.

With few exceptions, which will be discussed first, the K–Ar ages on 17 anorthoclase concentrates from 14 different pumice clasts collected from the KBS Tuff or its correlatives are remarkably concordant at ~1.9 Myr (Table 1).

Anorthoclase from sample KF43 yields anomalously old apparent ages of 4.11 and 7.46 Myr (Table 1). We attribute these poorly reproducible ages to the presence of variable but small amounts of old detrital K-feldspar in the aliquants used in the argon extractions. Careful petrographic examination of the mineral concentrate, however, did not lead to positive identification of detrital K-feldspar. Nevertheless, there is no doubt that old detrital material was being brought into the East Turkana Basin during deposition of the sediments. For example, we have obtained a K–Ar age of 456 ± 5 Myr ($K = 10.73\%$; radiogenic ⁴⁰Ar = 9.68×10^{-9} mol g⁻¹) on a large (30 mm) crystal of microcline (sample no. 79–457) found weathering out of sediments of the Koobi Fora Formation. Such an age is typical of the Mozambique Belt of East Africa^{26,27}. Contamination of the anorthoclase concentrate from KF43 by less than 0.5% of old K-feldspar of this kind would explain the aberrant results.

Two anorthoclase separates were obtained from pumice clast 7722–109C, one from the outer part of the clast (79–19), yielding an age of 2.00 ± 0.02 Myr, and the other (79–276) from the core of the clast, giving an age of 1.89 ± 0.02 Myr. These results clearly demonstrate that the outer layers of the pumice cobbles may contain detrital feldspar.

For two pumice clasts (KF38, KF47) two size fractions of feldspar were analysed, and in each case the finer fraction yields slightly older apparent ages. For KF47 the difference in measured age for the two size fractions is barely significant, but for KF38 the ages differ by 4.5%, and this difference is regarded as highly significant. This discrepancy may be because of incomplete extraction of the argon from the coarser fraction or because of minor contamination by detrital feldspar in the finer fraction. The latter explanation is preferred but cannot be proved.

Table 1 K–Ar ages on anorthoclase phenocrysts separated from pumice clasts found within the KBS Tuff or its correlatives, East Turkana, Kenya

Lab. no.	Field no.	Size fraction (μm)	K (wt %)	Wt used in Ar extraction (g)	Radiogenic ⁴⁰ Ar (10 ⁻¹¹ mol g ⁻¹)	100 Rad. ⁴⁰ Ar	Calculated age (Myr) ± 1 s.d.	Average age (Myr) ± 1 s.d.
						Total ⁴⁰ Ar		
Area 105								
78-1038	KF38	210-350	5.179, 5.147	2.664	1.700	48.6	1.90±0.02	1.89±0.02
				2.869	1.690	85.7	1.89±0.02	
				2.515	1.778	77.5	1.99±0.03	
78-1038	KF38	100-210	5.174, 5.128	2.606	1.762	40.9	1.97±0.03	1.98±0.02
Area 112								
79-17	7722-109A	125-350	5.108, 5.112	1.013	1.651	35.6	1.86±0.02	1.87±0.02
				1.020	1.662	37.5	1.87±0.02	
79-18	7722-109B	100-350	4.975, 4.949	0.904	1.622	41.1	1.88±0.02	1.88±0.02
79-19	7722-109C	100-350	5.163, 5.163	0.912	1.795	64.6	2.00±0.02	2.00±0.02
79-276	7722-109C	100-350	5.107, 5.112	1.931	1.682	84.4	1.90±0.03	1.89±0.02
				2.106	1.675	84.9	1.89±0.02	
				2.114	1.679	83.0	1.89±0.02	
79-277	7722-109D	180-350	5.100, 5.115	2.080	1.674	84.6	1.89±0.02	1.89±0.02
				2.028	1.636	68.3	1.87±0.02	
78-1039	KF39	210-420	5.042, 5.042	1.806	1.670	85.2	1.88±0.03	1.87±0.02
78-1040	KF40	180-350	5.100, 5.138	2.212	1.684	84.5	1.90±0.03	1.89±0.02
				1.979	1.655	86.7	1.87±0.02	
78-1041	KF41	180-350	5.097, 5.093	1.696	1.666	83.3	1.88±0.02	1.88±0.02
				1.859	6.556	87.9	7.46±0.08	
78-1043	KF43	210-420	5.061, 5.042	0.894	3.723	90.6	4.11±0.04	—
Area 131								
78-1044	KF44	100-500	5.136, 5.152	1.660	1.703	63.5	1.91±0.02	1.90±0.02
				2.239	1.687	82.2	1.89±0.02	
				2.399	1.673	78.5	1.88±0.02	
78-1047	KF47	210-350	5.138, 5.112	2.581	1.663	46.4	1.87±0.02	1.88±0.02
				2.528	1.693	53.8	1.90±0.03	
78-1047	KF47	100-210	5.117, 5.155	2.540	1.719	63.6	1.93±0.03	1.91±0.02
				2.576	1.681	72.9	1.89±0.02	
78-1048	KF48	210-350	5.127, 5.127	2.403	1.731	75.5	1.94±0.02	1.92±0.02
				1.013	1.671	55.5	1.87±0.02	
79-14	7722-107I	100-350	5.146, 5.171	1.022	1.700	55.2	1.90±0.02	1.88±0.02
				1.017	1.665	62.5	1.86±0.03	
79-15	7722-107II	100-350	5.129, 5.173	2.571	1.695	77.0	1.90±0.03	1.88±0.02

$$\lambda_e + \lambda_{ec} = 0.581 \times 10^{-10} \text{ yr}^{-1} \quad \lambda_\beta = 4.962 \times 10^{-10} \text{ yr}^{-1} \quad {}^{40}\text{K}/\text{K} = 1.167 \times 10^{-4} \text{ mol mol}^{-1}$$

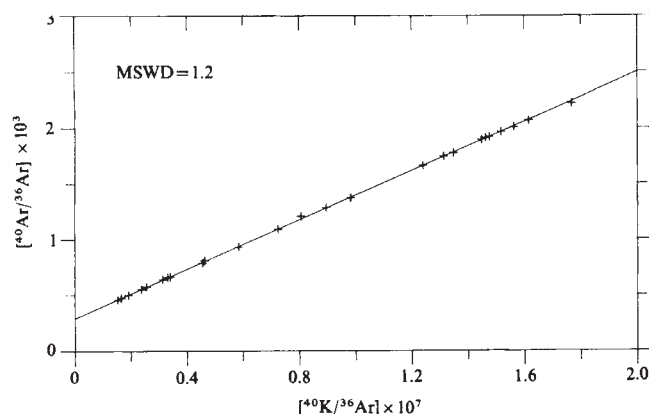


Fig. 2 K-Ar isochron diagram showing data from anorthoclase phenocrysts from pumice clasts found within the KBS Tuff. $R_1 = 293.8 \pm 2.4$; age = 1.89 ± 0.01 Myr.

Thus the K-Ar results from KF43, 79-19 and the finer fraction of KF38 are excluded from further consideration.

The average K-Ar ages for the other 14 anorthoclase separates from 13 different pumice clasts collected from the KBS Tuff, lie within the narrow range of 1.88–1.92 Myr, with a mean age of 1.89 ± 0.01 Myr (Table 1). Applying the isochron approach²¹ to these same data, yields a least squares fitted²⁸ line whose slope is equivalent to an age of 1.89 ± 0.01 Myr (Fig. 2), where the uncertainty quoted is at the 95% confidence level. Note that the line fits to within experimental error, as the value of the mean square of weighted deviates (MSWD) does not differ significantly from unity, and that the age is identical to that derived simply by averaging the K-Ar ages calculated for individual anorthoclase separates. In the isochron analysis no assumption is made concerning the $^{40}\text{Ar}/^{36}\text{Ar}$ ratio of the non-radiogenic component. Thus the fact that the value found (293.8 ± 2.4) is indistinguishable from that of atmospheric argon (295.5) provides independent verification that in the present case the non-radiogenic argon has atmospheric composition, as normally assumed in the calculation of K-Ar ages.

The concordancy of the K-Ar results argues strongly against the presence of contamination by old detrital K-feldspar, as it would be most unlikely that the proportion of contamination would be so uniform. Because all these samples consist of clear, fresh, high-temperature, volcanic anorthoclase, which normally has excellent retention properties for radiogenic argon, the measured age is interpreted as that of the eruptive episode that produced the pumice clasts and the enclosing tuff. The KBS Tuff has never been buried deeper than about 125 m, so that temperatures since deposition are not likely to have been raised much above ambient. We conclude, therefore, that loss of radiogenic argon from the anorthoclase phenocrysts by reheating subsequent to deposition of the pumice clasts, as proposed by Fitch and Miller¹⁰, is geologically unreasonable.

On the basis of chemical data on pumices, Cerling *et al.*²⁹ have argued that the tuff in area 112 (referred to as area 105-East in their paper) is not the same horizon as the KBS Tuff in areas 105 and 131. Our results on anorthoclase phenocrysts from the pumices in the three areas show no significant differences in potassium content or in measured K-Ar age. Thus our data strongly support the view of I.C. Findlater (personal communication) that the tuff in area 112 is stratigraphically equivalent to the type KBS Tuff in area 105.

Comparison with other results

The mean K-Ar age of 1.89 ± 0.01 Myr, obtained in the present study for anorthoclase separates from 13 pumice clasts in the KBS Tuff and its correlatives, is statistically indistinguishable from the mean K-Ar age of 1.83 ± 0.06 Myr calculated from the data of Drake *et al.*¹⁴, based on 22 measurements of feldspar and glass from 13 pumice clasts in the same horizon. Nevertheless,

the spread in apparent age for the samples analysed by Drake *et al.* is significantly greater than that found in the present study. Their reported ages for glass range from 1.73 ± 0.04 to 1.90 ± 0.03 Myr (six glass fractions from four samples), and for feldspar the ages range from 1.73 ± 0.03 to 1.92 ± 0.02 Myr (16 feldspar concentrates from 13 samples), a spread of some 10%, and much greater than the quoted errors of 1–2% for individual ages. It is significant that the oldest ages reported by Drake *et al.* are in excellent agreement with those found in the present study, and that the spread is towards younger ages.

Direct comparison between the results of Drake *et al.*¹⁴ and those determined in this laboratory is possible because measurements have been made in both laboratories on three pumice clasts which were collected jointly from the KBS Tuff by T. E. Cerling and A. J. W. G., and split into two. Separates for dating were prepared independently in the two laboratories, and the results are summarised in Table 2. The differences in measured potassium content for the feldspars are small and may simply reflect slight differences in the purity of the separates. However, the differences in reported radiogenic argon content range from 5.7 to 11%, with our results higher.

Table 2 Comparison of K-Ar ages obtained on separates from the same pumice clasts in the Berkeley and Canberra Laboratories

Berkeley no.	Canberra no.	Sample	K (wt%)	Radiogenic ^{40}Ar (10^{-11} mol g $^{-1}$)	Calculated age (Myr $\pm$ 1 s.d.)
KBS 77-109A		Feldspar	4.96	1.54	1.79 ± 0.02
	7722-109A	Feldspar	5.11	1.65	1.87 ± 0.02
		Glass	4.04	1.33	1.90 ± 0.03
KBS 77-109B		Feldspar	4.87	1.46	1.73 ± 0.03
	7722-109B	Feldspar	4.96	1.62	1.88 ± 0.02
KBS 77-109D		Feldspar	5.12	1.59	1.79 ± 0.02
	7722-109D	Feldspar	5.11	1.68	1.89 ± 0.02

Two points may be made. First, the Canberra results have a mean measured age of 1.88 ± 0.02 Myr for the three feldspars, whereas those measured in Berkeley by Drake *et al.*¹⁴ yield a mean of 1.77 ± 0.03 Myr, a difference of some 6%. Second, the age of volcanic glass from sample KBS 77-109B of 1.90 ± 0.03 Myr is concordant with the Canberra feldspar ages, but 9.8% greater than the age measured in Berkeley on feldspar from the same sample. Because glass normally is much more susceptible to loss of radiogenic argon than high temperature feldspar, it is indeed curious that the feldspar, as measured in Berkeley, yields a lower apparent age than the coexisting glass. A possible explanation for these results is that extraction of radiogenic argon from the feldspars was incomplete in the Berkeley experiments, leading to low apparent ages. We have mentioned previously the difficulty of extracting the argon quantitatively from feldspars of this kind, even when the samples have been melted in the vacuum system. Further intercomparisons are necessary to clarify this problem, but we wish to stress that overall the K-Ar age results from the two laboratories are in good agreement, and provide convincing evidence that the KBS Tuff has an age in the range 1.8–1.9 Myr.

It now seems clear that the earlier K-Ar and $^{40}\text{Ar}/^{39}\text{Ar}$ total fusion results on feldspar from pumice clasts of the KBS Tuff reported by Fitch and Miller^{9,10}, and Fitch *et al.*¹², must be regarded as imprecise, as is evident from the large scatter of their data. This imprecision arises partly from the high atmospheric argon contamination found in most of their argon analyses, generally in the range 80–95% of the total ^{40}Ar . The meaning of the complex $^{40}\text{Ar}/^{39}\text{Ar}$ age spectra found^{9,10,12} for anorthoclase remains unclear. Three of the samples yield apparent ages for part of the age spectra between 2.4 and 2.5 Myr, possibly reflecting some detrital contamination. Fitch and Miller¹⁰ interpret apparent ages of around 1 Myr and between 1.8 and 2 Myr in some of the age spectra as indicating overprinting with loss of radiogenic argon at these times. On the basis of our data we believe the apparent ages of 1.8–2 Myr are recording the time of crystallisation and cooling of the feldspars, rather than

overprinting. It has been noted that the postulated overprinting events seem unreasonable on geological grounds. Indeed the presence of fresh glass in some of the pumice clasts, together with the relatively good agreement between the K-Ar ages on the glass and coexisting feldspar, as reported by Drake *et al.*¹⁴, provides powerful evidence against such overprinting.

Finally, the recently re-determined fission track ages, averaging 1.87 ± 0.04 Myr, on zircon separated from pumice clasts in the KBS Tuff, reported by Gleadow²⁷, are in excellent agreement with our K-Ar data, and resolve an outstanding problem.

Conclusions

The mean K-Ar age of 1.89 ± 0.01 Myr found in the present study for anorthoclase separates from 13 pumice clasts of the KBS Tuff is interpreted as the age of crystallisation and cooling of the pumices, which were deposited shortly thereafter in the East Turkana Basin. We believe that these data provide the best estimate presently available for the age of this stratigraphically important horizon within the East Turkana sedimentary

sequence. This view is reinforced by the satisfactory agreement between our results and those obtained by Drake *et al.*¹⁴ in the Berkeley laboratory on anorthoclase and glass, and the fission track ages on zircon given by Gleadow²⁷. The combined results seem to resolve the long-standing controversy surrounding the age of the KBS Tuff and indicate that this horizon was deposited in the late Pliocene as the Plio-Pleistocene boundary has an estimated age of ~ 1.6 – 1.7 Myr (ref. 30). Acceptance of such an age for the KBS Tuff also resolves the correlation problems, recently summarised by Behrensmeyer³¹ and Brown *et al.*³², between the sequences in East Turkana and the Omo area of southern Ethiopia. The present results also suggest that the KBS Tuff was deposited during the time interval represented by the Olduvai normal polarity event in the Matuyama reversed epoch of the geomagnetic polarity time scale^{33–36}.

We thank G.L.I. Isaac and R. E. F. Leakey, joint leaders of the Koobi Fora Research Project, for their encouragement. Support in the field by the Koobi Fora Research Project is gratefully acknowledged. We thank I.C. Findlater for his guidance in the field. Sample preparation and mineral separations were undertaken by M. Cowan and R. Rudowski.

Received 29 October 1979; accepted 30 January 1980.

- Coppens, Y., Howell, F. C., Isaac, G. L. I. & Leakey, R. E. F. (eds) *Earliest Man and Environments in the Lake Rudolf Basin* (University of Chicago Press, 1976).
- Bishop, W. W. (ed.) *Geological Background to Fossil Man* (Scottish Academic, Edinburgh, 1978).
- Leakey, M. G. & Leakey, R. E. F. (eds) *Koobi Fora Research Project Vol. 1* (Clarendon, Oxford, 1978).
- Behrensmeyer, A. K. *Nature* **226**, 225–226 (1970).
- Vondra, C. F., Johnson, G. D., Bowen, B. E. & Behrensmeyer, A. K. *Nature* **231**, 245–248 (1971).
- Bowen, B. E. & Vondra, C. F. *Nature* **242**, 391–393 (1973).
- Leakey, R. E. F. *Nature* **226**, 223–224 (1970).
- Leakey, M. D. *Nature* **226**, 228–230 (1970).
- Fitch, F. J. & Miller, J. A. *Nature* **226**, 226–228 (1970).
- Fitch, F. J. & Miller, J. A. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y., Howell, F. C., Isaac, G. L. I. & Leakey, R. E. F.) 123–147 (University of Chicago Press, 1976).
- Curtis, G. H., Drake, R., Cerling, T. & Hampel, J. *Nature* **258**, 395–398 (1975).
- Fitch, F. J., Hooker, P. J. & Miller, J. A. *Nature* **263**, 740–744 (1976).
- Hurford, A. J., Gleadow, A. J. W. & Naeser, C. W. *Nature* **263**, 738–740 (1976).
- Drake, R. E., Curtis, G. H., Cerling, T. E., Cerling, B. W. & Hampel, J. *Nature* **283**, 368–372 (1980).
- Findlater, I. C. in *Koobi Fora Research Project Vol. 1* (eds Leakey, M. G. & Leakey, R. E. F.) 14–31 (Clarendon, Oxford, 1978).
- Vondra, C. F. & Bowen, B. E. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 395–414 (Scottish Academic, Edinburgh, 1978).
- Findlater, I. C. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y. *et al.*) 94–104 (University of Chicago Press, 1976).
- Findlater, I. C. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 415–420 (Scottish Academic, Edinburgh, 1978).
- Cooper, J. A. *Geochim. cosmochim. Acta* **27**, 525–546 (1963).
- Webb, A. W. & McDougall, I. *Earth planet. Sci. Lett.* **3**, 41–47 (1967).
- McDougall, I., Polach, H. A. & Stipp, J. J. *Geochim. cosmochim. Acta* **33**, 1485–1520 (1969).
- McDougall, I. & Schmincke, H.-U. *Bull. Volcanol.* **40**, 57–77 (1977).
- Steiger, R. H. & Jäger, E. *Earth planet. Sci. Lett.* **36**, 359–362 (1977).
- Wright, T. L. *Am. Miner.* **53**, 38–87 (1968).
- Wright, T. L. & Stewart, D. B. *Am. Miner.* **53**, 88–104 (1968).
- Cahen, L. & Snelling, N. J. *The Geochronology of Equatorial Africa* (North-Holland, Amsterdam, 1966).
- Gleadow, A. J. W. *Nature* **284**, 225–230 (1980).
- York, D. *Earth planet. Sci. Lett.* **5**, 320–324 (1969).
- Cerling, T. E., Brown, F. H., Cerling, B. W., Curtis, G. H. & Drake, R. E. *Nature* **279**, 118–121 (1979).
- Haq, B. U., Berggren, W. A. & Van Couvering, J. A. *Nature* **269**, 483–488 (1977).
- Behrensmeyer, A. K. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 421–440 (Scottish Academic, Edinburgh, 1978).
- Brown, F. H., Howell, F. C. & Eck, G. G. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 473–498 (Scottish Academic, Edinburgh, 1978).
- Brook, A. & Isaac, G. L. I. *Nature* **247**, 344–348 (1974).
- Hillhouse, J. W., Ndombi, J. W. M., Cox, A. & Brock, A. *Nature* **265**, 411–415 (1977).
- Brook, A. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 471 (Scottish Academic, Edinburgh, 1978).
- McDougall, I. in *The Earth: Its Origin, Structure and Evolution* (ed. McElhinny, M. W.) 534–566 (Academic, London, 1979).

Chicken lens crystallin DNA sequences show at least two δ -crystallin genes

Suraj P. Bhat, Raymond E. Jones, Margery A. Sullivan & Joram Piatigorsky

Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20205

Restriction analysis of chicken DNA using probes derived from a cloned cDNA and analysis of cloned genomic DNA fragments containing δ -crystallin gene sequences have indicated the presence of at least two non-allelic genes for δ -crystallin, the first and principal crystallin synthesised in the embryonic chicken lens. Electron microscopic analyses of three cloned genomic fragments revealed that the δ -crystallin mRNA gene sequences are interrupted at least 14 times in one of the δ -crystallin genes.

CRYSTALLINS are a group of structural proteins which constitute the bulk of the protein in the vertebrate lens¹. Since they are tissue specific, the crystallins are advantageous markers for cellular differentiation^{2–5}. δ -Crystallin, found only in birds and reptiles^{6–8}, is a particularly useful index of lens cell differentiation. It is the first crystallin to appear after lens induction in the chicken embryo^{9–11}. Further differentiation of lens fibre cells is correlated with a rapid accumulation of δ -crystallin mRNA^{12–14}. δ -Crystallin represents 70 to 80% of the protein synthesised by the embryonic lens fibre cells which, by 15 days of

development, contain on average $1\text{--}2 \times 10^5$ molecules of δ -crystallin mRNA per cell¹⁵. The embryonic chicken lens fibre cells synthesise much smaller amounts of α - and β -crystallin¹⁶. After hatching, δ -crystallin synthesis decreases and β -crystallin synthesis predominates¹⁶. δ -Crystallin is composed of at least two similar polypeptides^{17,18}. Although δ -crystallin mRNA has not been fractionated into more than one species^{13,19}, it is possible that the δ -crystallin polypeptides may be encoded by similar but non-identical mRNAs¹⁵.

An understanding of the regulation of δ -crystallin synthesis

during development will ultimately require knowledge of the structure of the δ -crystallin gene(s). Annealing experiments with cDNA have shown that the δ -crystallin gene sequences are situated in the non-reiterated region of the genome²⁰, but the actual number and organisation of δ -crystallin gene(s) have not been established. We have recently cloned δ -crystallin cDNA in a bacterial plasmid²¹, which provides an opportunity for analysis of the natural δ -crystallin gene(s). In this article we have used this recombinant plasmid to investigate the organisation of the δ -crystallin sequences in chicken DNA and in genomic fragments cloned from it.

Identification and cloning of δ -crystallin gene sequences

High molecular weight DNA was extracted from erythrocytes of White Leghorn chickens (gs⁻, Spafas Incorporated). The DNA was limit-digested with *Eco*RI and chromatographed on an RPC-5 column followed by agarose gel electrophoresis²². The fragments containing δ -crystallin sequences were identified by Southern blotting²³ and hybridisation to δ -crystallin ³²P-cDNA (ref. 21), synthesised from sucrose density gradient purified mRNA¹⁹. Four principal bands of DNA hybridised to the ³²P-cDNA (data not shown). Four major *Eco*RI fragments containing δ -crystallin sequences were also found with DNA isolated from pooled embryos or from adult liver or cultured embryonic fibroblasts derived from individual chickens. In these experiments nick-translated p δ Cr-2 DNA was used as a probe. p δ Cr-2 is a recombinant plasmid containing approximately 1.3 kilobases of δ -crystallin cDNA cloned in the bacterial plasmid pBR322 (ref. 21). δ -Crystallin mRNA is approximately 2 kilobases long^{19,21}. p δ Cr-2 DNA was used as a probe for all subsequent experiments.

The fragments were purified from the appropriate RPC-5 column fractions of the erythrocyte DNA by preparative agarose gel electrophoresis²⁴ and identified by analytical agarose gel electrophoresis, Southern blotting and hybridisation to ³²P-cDNA as above. We used the *in vitro* packaging method of Blattner *et al.*²⁵ to clone the two largest fragments in bacteriophage λ Charon 4A and the third largest fragment in bacteriophage λ gtWES (ref. 26). Charon 4A was selected because the lengths of the two biggest fragments were initially thought to be larger than 10 kilobases. Re-electrophoresis of the purified fragments containing δ -crystallin sequences indicated that their lengths were approximately 12, 7.5, 4 and 3 kilobases. The recombinant clones were identified by the *in situ* screening procedure²⁷. We refer to the cloned 7.5-, 12- and 4-kilobase fragments as g δ Cr-1, g δ Cr-2 and g δ Cr-3 respectively (for genomic δ -crystallin clones 1, 2 and 3).

To determine the relative orientation of the 7.5-, 12- and 4-kilobase fragments in the genome, we took advantage of an *Eco*RI site within the cloned cDNA insert in p δ Cr-2 (ref. 21). Digestion of p δ Cr-2 with *Eco*RI generates one fragment containing approximately 550 base pairs from the 3' half of the cloned δ -crystallin cDNA and another fragment containing approximately 700 base pairs from the 5' half of the cloned cDNA. The extreme 5' sequences of the δ -crystallin mRNA are not represented in p δ Cr-2. The 5' probe derived from p δ Cr-2 hybridised strongly to the cloned 7.5-kilobase fragment, extremely weakly to the 12-kilobase fragment and not at all to the 4-kilobase fragment. The 3' probe hybridised exclusively to the 12- and 4-kilobase cloned fragments. We conclude from this experiment that the 7.5-kilobase fragment contains δ -crystallin sequences located 5' to those in the 12- or 4-kilobase fragments with respect to δ -crystallin mRNA.

Visualisation of δ -crystallin sequences in the genomic clones

Electron microscopy was used to investigate the organisation of the sequences complementary to δ -crystallin mRNA within the cloned fragments. Heteroduplexes were formed between the DNAs of g δ Cr-1 and g δ Cr-2, and g δ Cr-1 and g δ Cr-3;

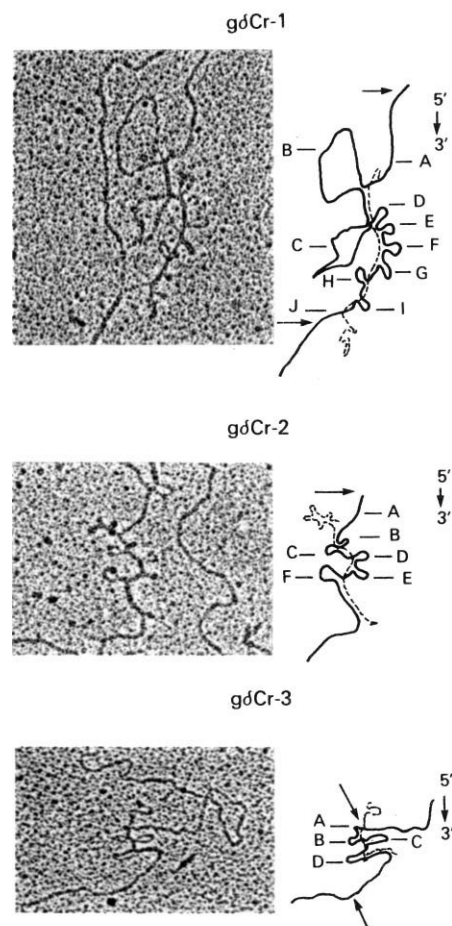


Fig. 1 Electron micrographs of the hybrids formed between δ -crystallin mRNA and the 7.5-kilobase fragment in g δ Cr-1, the 12-kilobase fragment in g δ Cr-2 and the 4-kilobase fragment in g δ Cr-3, respectively. Heteroduplexes were formed between g δ Cr-1 and g δ Cr-2 or g δ Cr-1 and g δ Cr-3 DNA (10 μ g ml⁻¹ each) in 70% formamide, 0.1M Tris-HCL, pH 7.5, 0.15M NaCl, 0.01M EDTA and 25 μ g ml⁻¹ of δ -crystallin mRNA. The cloned DNAs were first kept at 85 °C for 10 min and then the mRNA was added to the reaction mixture. The hybridisation was continued at 43 °C for 2 h. Reaction mixtures were diluted 1:20 in the hybridisation buffer, prepared for electron microscopy²¹ and examined in a Philips 300 electron microscope. An illustrative tracing is presented for each electron micrograph. The intervening sequences (IVS) in each molecule are indicated by letters. IVS B in g δ Cr-2 was observed in only one-third of the molecules. This may be due to the short hybridised region between A and B. Alternatively, B may be a spreading artefact. It is even possible that its occasional presence is due to the existence of two different δ -crystallin mRNAs. Digestion of g δ Cr-1 by *Eco*RI released two cloned fragments of about 8.0 kilobases indicating end-to-end ligation of two fragments. This explains the viability of the 7.5-kilobase fragment in λ Charon 4A, which requires at least 10 kilobases of inserted DNA. Only one of these fragments (7.5 kilobases long) contained δ -crystallin sequences. Similarly, two cloned fragments of 5.0 kilobases and 4.0 kilobases were released when g δ Cr-3 was digested by *Eco*RI. Out of these two fragments only the 4-kilobase fragment contained δ -crystallin sequences. Arrows indicate the *Eco*RI sites: —: single-stranded DNA; --- mRNA.

δ -crystallin mRNA was then hybridised to the single-stranded region of each of the heteroduplexes. The upper panel of Fig. 1 shows an electron micrograph of a hybrid between δ -crystallin mRNA and the complementary sequences in g δ Cr-1 DNA. The RNA-DNA hybrids were interrupted by eight DNA loops, indicating discontinuity of the sequences complementary to δ -crystallin mRNA. These intervening sequences (IVS) are denoted by letters A-J; the hybridised regions are denoted by numbers 1-9 (see Fig. 2). The cumulative length of the hybri-

dised regions, which we call mRNA gene sequences (mRGS), was 1.4 ± 0.4 kilobases (mean $\pm$ s.d.); the length of the intervening sequences was 6.2 ± 0.9 kilobases. The long RNA tail is the 3' part of δ -crystallin mRNA, since more than 60% of the mRNA sequences hybridised to g δ Cr-1, and g δ Cr-1 did not hybridise to the 3' half of the cloned δ -crystallin cDNA. The non-hybridised stretch of DNA-labelled A may be an intervening sequence because a short tail of RNA was occasionally observed to the left of mRGS 1. This suggests that a δ -crystallin mRNA gene sequence is missing from the 5' end of the 7.5-kilobase fragment. Assuming that the sequences at the extreme 5' end of δ -crystallin mRNA are not present in g δ Cr-1, IVS A would be the largest intervening sequence (at least 1.5 kilobases) in g δ Cr-1. It is interesting that the intervening sequence separating the 45-base pair leader sequence of ovalbumin mRNA in the chicken genome is 1.6 kilobases long²⁸.

The electron micrograph in Fig. 1 (middle) shows a hybrid between δ -crystallin mRNA and the single-stranded DNA in g δ Cr-2. The prominent RNA tail seen in this hybrid is the 5' end of δ -crystallin mRNA, since only the 3' half of the cloned δ -crystallin cDNA hybridises well to this fragment. The small knob-like structure present at the 3' end of the hybrid probably represents the poly(A) tail of the mRNA. Six intervening sequences (A-F) were observed in this hybrid. The cumulative length of the intervening sequences in g δ Cr-2 was 1.9 ± 0.3 kilobases and that of the mRNA gene sequences was 0.8 ± 0.2 kilobases.

The electron micrograph in Fig. 1 (lower) shows a hybrid between δ -crystallin mRNA and the single-stranded DNA in g δ Cr-3. The δ -crystallin sequences within this cloned 4-kilobase fragment hybridised only to the 3' half of the cloned cDNA probe. The long RNA tail is thus the 5' side of the δ -crystallin mRNA. Four intervening sequences (A-D) interrupted the δ -crystallin mRNA gene sequences in this cloned fragment. By analogy with g δ Cr-2, there may be a fifth intervening sequence to the left of the mRNA gene sequence 1 (see Fig. 2) in g δ Cr-3. This cannot, however, be established by electron microscopy. The cumulative length of the intervening sequences of the cloned 4-kilobase fragment is about 1.7 ± 0.3 kilobases, and that of the mRNA gene sequences is about 0.6 ± 0.1 kilobases—values in close agreement with the corresponding values for the 12-kilobase fragment. The 12- and 4-kilobase fragments thus appear to contain very similar but not identical 3' halves of separate δ -crystallin genes (see below).

Diagrammatic representations of the δ -crystallin gene sequences in the three cloned fragments are shown in Fig. 2. δ -Crystallin DNA sequences in g δ Cr-1 must be associated with those in g δ Cr-2 or g δ Cr-3 to form one of the natural δ -crystallin genes. Either combination contains a sufficient number of δ -crystallin mRNA gene sequences to account for a δ -crystallin gene, except probably the 5' leader sequence. The cumulative length of the mRNA gene sequences in the 7.5 (g δ Cr-1) and the 12 (g δ Cr-2) kilobase fragments is 2.2 ± 0.6 kilobases, and the cumulative length of these sequences in the 7.5 and the 4 (g δ Cr-3) kilobase fragments is 2.0 ± 0.5 kilobases. (δ -Crystallin mRNA is approximately 2.0 kilobases long^{19,21}.) Since the total length of g δ Cr-1 and g δ Cr-2 or g δ Cr-3 is about 10 kilobases, the intervening sequences amount to approximately 80% of the δ -crystallin gene.

The present data suggest that the two δ -crystallin genes are very similar, as judged by the pattern of intervening sequences in both 3' halves of the genes. The fact that the two 3' halves of the δ -crystallin genes are contained in a 4- and 12-kilobase fragment after *Eco*RI digestion indicates that the two genes have different flanking sequences at their 3' ends. It is possible that the 5' halves of both genes are present in the *Eco*RI fragment of approximate length 7.5 kilobases. This is consistent with the results obtained on the quantitation of hybridisation of ³²P-cDNA to different *Eco*RI fragments separated by electrophoresis after reverse-phase column chromatography. The relative amount of ³²P-cDNA hybridised to each δ -crystallin DNA band was estimated by scanning the autoradiogram. If one takes the amount of radioactivity hybridised to the 3-kilobase fragment as one unit, then the relative amounts of ³²P-cDNA hybridised to the 4-, 7.5- and 12-kilobase fragments were 2.1, 7.0 and 2.3 units, respectively. Thus, there was almost four times as much hybridisation to the 7.5-kilobase fragment as to either the 4- or 12-kilobase fragments. The remaining 3.0-kilobase fragment derived by *Eco*RI digestion of the chicken DNA hybridised much more weakly to the ³²P-cDNA, suggesting that it contains relatively few δ -crystallin mRNA gene sequences.

Further evidence for two δ -crystallin genes

In contrast to the results obtained with *Eco*RI, *Bam*HI digestion of chicken DNA gave two major fragments of approximately 8 and 10 kilobases which hybridised to nick-translated p δ Cr-2 DNA (Fig. 3). This is of interest since the cloned cDNA has a central *Eco*RI site but lacks a site for *Bam*HI²¹. Thus, the two

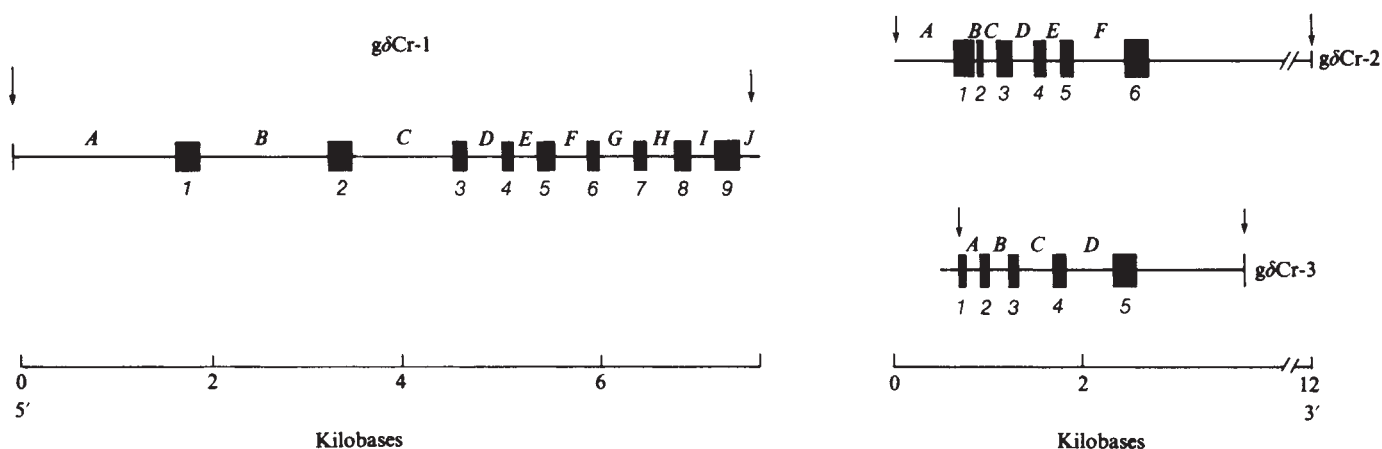


Fig. 2 Diagrammatic representation of intervening (IVS) and mRNA gene sequences (mRGS) in the cloned δ -crystallin gene fragments. Hybrid molecules of g δ Cr-1 DNA: δ -crystallin mRNA, g δ Cr-2: δ -crystallin mRNA, and g δ Cr-3: δ -crystallin mRNA (see Fig. 1) were measured, using a Numonics digitiser and statistically analysed. The orientation of the fragments was determined as given in the text. The arrows ($\downarrow$) indicate the *Eco*RI sites. The measured lengths ($\pm$ s.d.) of IVS (letters) and mRGS (numbers) were: g δ Cr-1 (36 molecules): A, 1.60 ± 0.3 ; B, 1.30 ± 0.10 ; C, 1.10 ± 0.16 ; D, 0.40 ± 0.07 ; E, 0.30 ± 0.05 ; F, 0.39 ± 0.06 ; G, 0.32 ± 0.05 ; H, 0.35 ± 0.06 ; I, 0.23 ± 0.05 ; J, 0.19 ± 0.05 ; 1, 0.24 ± 0.09 ; 2, 0.24 ± 0.05 ; 3, 0.13 ± 0.03 ; 4, 0.09 ± 0.03 ; 5, 0.12 ± 0.02 ; 6, 0.11 ± 0.03 ; 7, 0.17 ± 0.04 ; 8, 0.07 ± 0.03 ; 9, 0.19 ± 0.06 ; g δ Cr-2 (24 molecules): A, 0.67 ± 0.09 ; B, not measured; C, 0.28 ± 0.05 ; D, 0.23 ± 0.04 ; E, 0.23 ± 0.04 ; F, 0.55 ± 0.1 ; 1, 0.18 ± 0.04 ; 2, not measured; 3, 0.13 ± 0.04 ; 4, 0.10 ± 0.03 ; 5, 0.17 ± 0.03 ; 6, 0.24 ± 0.06 ; g δ Cr-3 (23 molecules): A, 0.13 ± 0.03 ; B, 0.25 ± 0.06 ; C, 0.39 ± 0.06 ; D, 0.50 ± 0.13 ; 1, 0.06 ± 0.02 ; 2, 0.08 ± 0.02 ; 3, 0.10 ± 0.03 ; 4, 13 ± 0.3 ; 5, 0.24 ± 0.05 . —, IVS; ■, mRGS. The segment shown as a broken line in g δ Cr-3 was not visualised but may be present by analogy with g δ Cr-2

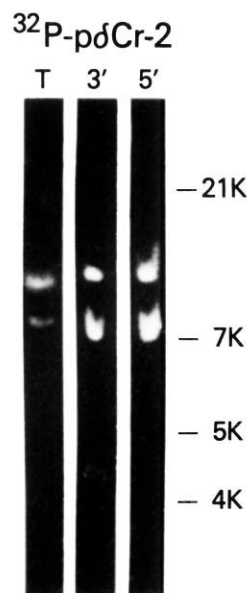


Fig. 3 Autoradiogram of a Southern blot of *Bam*HI-digested chicken DNA after hybridisation to 32 P-labelled p δ Cr-2 DNA. DNA isolated from 6-day-old chicken embryos was digested with *Bam*HI in the buffer recommended by the manufacturers (New England Biolabs), subjected to electrophoresis in 1% agarose gel, transferred to Millipore filters²³ and hybridised⁵⁷ to different 32 P-labelled probes²¹. Each sample (20 μ g) was digested with 40 enzyme units for 12 or 24 h. Digestion with addition of fresh enzyme at 2, 4 and 6 h (final substrate: enzyme ratio 1:6) did not change the hybridisation pattern. p δ Cr-2 was cut by *Eco*RI to give two fragments 4.2 kilobases and 1.3 kilobases long containing 5' and 3' sequences of the cloned cDNA insert, respectively. These two fragments were separated by electrophoresis in a 5% polyacrylamide gel and recovered by electrophoretic elution⁵⁸, concentrated by DEAE-cellulose chromatography and ethanol precipitated. Nick translation was done with 32 P-dCTP (300 Ci mmol⁻¹) according to Maniatis *et al.*⁵⁹. Each hybridisation reaction (20 ml) contained 4–6 $\times 10^6$ c.p.m. (10^7 – 10^8 c.p.m. μ g⁻¹). T = total (uncut nick-translated p δ Cr-2 DNA).

*Bam*HI fragments were not generated by cutting in the genomic regions corresponding to the cloned cDNA sequences. Unlike the cloned *Eco*RI fragments, each of the *Bam*HI fragments contained sequences complementary to the 3' and 5' halves of the cloned cDNA probe (Fig. 3). This suggests that the *Bam*HI fragment was not produced by cutting an intervening sequence within the δ -crystallin gene. The presence of two *Bam*HI fragments is consistent with the possibility that there are two separate genes for δ -crystallin. The four *Eco*RI fragments or the two *Bam*HI fragments containing δ -crystallin sequences are apparently not derived from two structural alleles of the δ -crystallin gene, since similar restriction fragments were obtained from DNA of inbred chickens (line 6, subline 3, US Dept of Agriculture, East Lansing, Michigan) and from DNA of a parthenogenetic turkey²⁹ (data not shown).

Discussion

In this investigation we have isolated, cloned and determined the pattern of intervening sequences in three δ -crystallin gene fragments produced by *Eco*RI digestion of the chicken genome. This is the first time that gene sequences for a lens crystallin gene have been examined. δ -Crystallin gene sequences are of special interest because they are expressed principally in the embryo^{3,4,10,16} and their activity can be studied *in vivo*^{4–11} and in tissue culture^{12,15}. The present study thus provides a starting point for further analysis of the regulation of δ -crystallin gene expression at the molecular level.

Our data are consistent with the existence of at least two separate non-allelic δ -crystallin genes in chickens. The fact that

*Bam*HI DNAs digested from extensively inbred chickens and a parthenogenetic turkey have similar restriction patterns for the δ -crystallin sequences to the *Bam*HI-treated DNA from the commercially available chickens (Truslow Farms) argues against interpreting our data as evidence for two structural alleles of one δ -crystallin gene. The similarity in the restriction patterns of δ -crystallin DNA sequences in turkeys with those in chickens is consistent with the similarity in the δ -crystallin protein in these two species⁸. The two δ -crystallin genes could be adjacent with a relatively small intergenic spacer, as in the human α -globin genes^{30,31}, or they may be widely separated in the genome. The data do not exclude the presence of more than two identical δ -crystallin genes which are tandemly repeated, but previous hybridisation kinetic experiments with cDNA do not support this notion²⁰.

δ -Crystallin is composed of subunits with approximate molecular weights of 50,000 and 48,000 (ref. 17). Peptide analyses have indicated that these subunits must have very similar primary structures¹⁷; only two tryptic peptide differences have been noted between the δ -crystallin subunits³². The presence of two different genes for δ -crystallin suggests that the larger and smaller subunits of this protein may be synthesised on separate mRNAs. This would be of considerable interest with respect to the control of their translation, since it is known that the ratio of synthesis of the δ -crystallin subunits is affected by the intracellular concentration of electrolytes³³. Interestingly, the rat preproinsulin gene has recently been shown to be encoded by two similar, non-allelic genes which are equally expressed^{11,34}. However, the possible existence of two genes for δ -crystallin does not necessarily mean that each subunit is encoded by a separate gene. One subunit may be derived from the other by post-translational processing³⁵, or differential splicing of transcripts from a single gene could lead to different mRNAs^{36–38}.

The present data demonstrate that the δ -crystallin genes are highly interrupted in comparison with other eukaryotic genes^{39–49}. The chicken conalbumin gene which is interrupted 17 times⁵⁰ is the only other gene reported yet that has as many or more intervening sequences as the δ -crystallin gene. If the mRNA gene sequences are separated into functional units or domains^{51,52} the large number of intervening sequences may indicate that δ -crystallin is structurally complex. Some interesting features of δ -crystallin are that it has a greater content of leucine and isoleucine than other eukaryotic structural proteins⁵³, it has a very low proportion of aromatic amino acids⁵³, and it has a high degree of α -helical structure^{54,55}. Finally, it would be particularly interesting if the unusual abundance of intervening sequences in the δ -crystallin gene was related to its distribution among animals. δ -Crystallin is confined to birds and reptiles, as judged by immunological^{6,7} and cDNA hybridisation⁸ studies. This special distribution of δ -crystallin differs from the ubiquitous distribution of other proteins, such as actin, histones or cytochrome c, for example, which have uninterrupted genes⁵⁶.

We thank Ann Skalka, Shirley M. Tilghman and Marion Nau for supplying the initial high molecular weight DNA and the RPC-5 column fractions, Dr E. Buss for a parthenogenetic turkey, Jonathan G. Seidman and Philip Leder for helpful discussions and advice, Jacob V. Maizel Jr for providing facilities for the electron microscopy and Terri Broderick for typing. The recombinant DNA experiments were performed in a P3 laboratory in accordance with the NIH Guidelines.

Received 27 November 1979; accepted 18 January 1980.

1. Harding, J. J. & Dilley, K. J. *Expl Eye Res.* **22**, 1–73 (1976).
2. Papaconstantinou, J. *Science* **156**, 338–346 (1967).
3. Clayton, R. M. *Curr. Topics dev. Biol.* **5**, 115–180 (1970).
4. Piatigorsky, J. *dev. Biol.* **40**, f21–f23 (1974).
5. Bloemendal, H. *Science* **197**, 127–138 (1977).
6. Manski, W. & Halbert, S. P. in *Proc. 12th Colloq. Bruges* (ed. Peeters, H.) 117–134 (Elsevier, Amsterdam, 1964).
7. Clayton, R. M. in *The Eye* Vol. 5 (eds Davson, H. & Graham, L. T.) 399–494 (Academic, London, 1974).
8. Williams, L. A. & Piatigorsky, J. *Eur. J. Biochem.* **100**, 349–357 (1979).
9. Rabaey, M. *Expl Eye Res.* **1**, 310–316 (1962).

10. Zwaan, J. & Ikeda, A. *Expl Eye Res.* **7**, 301–311 (1968).
11. Shinohara, T. & Piatigorsky, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2808–2812 (1976).
12. Milstone, L. M., Zelenka, P. & Piatigorsky, J. *Devl Biol.* **48**, 197–204 (1976).
13. Zelenka, P. & Piatigorsky, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1896–1900 (1974).
14. Jackson, J. F. *et al. Devl Biol.* **65**, 383–395 (1978).
15. Piatigorsky, J., Beebe, D. C., Zelenka, P., Milstone, L. M. & Shinohara, T. *INSERM* **60**, 85–112 (1976).
16. Genis-Galvez, J. M., Maisel, H. & Castro, J. *Expl Eye Res.* **7**, 593–602 (1968).
17. Reszelbach, R., Shinohara, T. & Piatigorsky, J. *Expl Eye Res.* **25**, 583–593 (1977).
18. Thomson, I., Wilkinson, C. E., Burns, A. T. H., Truman, D. E. S. & Clayton, R. M. *Expl Eye Res.* **26**, 351–362 (1978).
19. Zelenka, P. & Piatigorsky, J. *Expl Eye Res.* **22**, 115–124 (1976).
20. Zelenka, P. & Piatigorsky, J. *J. biol. Chem.* **252**, 4294–4298 (1976).
21. Bhat, S. P. & Piatigorsky, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3299–3303 (1979).
22. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 4406–4410 (1977).
23. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
24. Polsky, F., Edgell, M. H., Seidman, J. G. & Leder, P. *Analyt. Biochem.* **87**, 397–410 (1978).
25. Blattner, F. R. *et al. Science* **202**, 1279–1284 (1978).
26. Leder, P., Tiemeier, D. & Enquist, L. *Science* **196**, 175–177 (1977).
27. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
28. Dugaiczky, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2253–2257 (1979).
29. Darcey, K. M., Buss, E. G., Bloom, S. E. & Olsen, M. W. *Genetics* **69**, 479–489 (1971).
30. Orkin, S. H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5950–5954 (1978).
31. Little, P. F. R., Flavell, R. A., Kooter, J. M., Anisson, G. & Williamson, R. *Nature* **278**, 227–231 (1979).
32. Shinohara, T., Reszelbach, R. & Piatigorsky, J. *Expl Eye Res.* (in the press).
33. Shinohara, T. & Piatigorsky, J. *Nature* **270**, 406–411 (1977).
34. Lomedico, P. *et al. Cell* **18**, 545–558 (1979).
35. Piatigorsky, J. *Curr. Topics Eye Res.* **3** (eds Davson, H. & Zaduanisky, J.) (Academic, New York, in the press).
36. Chow, L. T. & Broker, T. R. *Cell* **15**, 497–510 (1978).
37. Crick, F. *Science* **204**, 264–271 (1979).
38. Flint, S. J. *Am. Scientist* **67**, 300–311 (1979).
39. Gannon, F. *et al. Nature* **278**, 428–434 (1979).
40. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725–729 (1978).
41. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976).
42. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11–17 (1978).
43. Rabbitts, T. H. & Forster, A. *Cell* **13**, 319–327 (1978).
44. Nguyen-Huu, M. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 76–80 (1979).
45. Royal, A. *et al. Nature* **279**, 125–132 (1979).
46. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097–1108 (1977).
47. Catterall, J. F. *et al. Nature* **278**, 323–327 (1979).
48. Sargent, T. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 3256–3260 (1979).
49. Smithies, O. *et al. Science* **202**, 1284–1289 (1978).
50. Cochet, M. *Nature* **282**, 567–574 (1979).
51. Sakano, H. *et al. Nature* **277**, 631–633 (1979).
52. Blake, C. C. F. *Nature* **277**, 598 (1979).
53. Piatigorsky, J., Zelenka, P. & Simpson, R. T. *Expl Eye Res.* **18**, 435–446 (1974).
54. Piatigorsky, J., Horwitz, J. & Simpson, R. T. *Biochim. biophys. Acta* **490**, 279–289 (1977).
55. Yu, N.-T., East, E. J., Chang, R. C. C. & Kuck, J. F. R. *J. Expl Eye Res.* **24**, 321–334 (1977).
56. Dawid, I. B. & Wahli, W. *Devl Biol.* **69**, 305–328 (1979).
57. Denhardt, D. T. *Biochem. biophys. Res. Commun.* **23**, 641–646 (1966).
58. McDonnell, M. W., Simon, M. N. & Studier, F. W. *J. molec. Biol.* **110**, 119–146 (1977).
59. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).

Contribution of immunoglobulin heavy-chain variable-region genes to antibody diversity

T. H. Rabbitts, G. Matthyssens* & P. H. Hamlyn

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

A mouse cloned cDNA probe containing a variable (V) region belonging to the V_HIII subgroup has been used in filter hybridisations to estimate the number of heavy-chain V-genes in this subgroup of mouse and human DNA. There seem to be about 10 and 20 V_H-genes hybridising to this probe in mouse and human DNA, respectively. Studies of cross-hybridisation of the related V_κ-genes from MOPC21 and MPC11 myelomas indicate that the experiments detect all members of the V_HIII subgroup.

THE antibody combining site comprises areas of the variable (V) regions of both heavy (H) and light (L) chains so that antibody diversity is a function of the germ-line diversity of both H- and L-chain genes combined with possible somatic modifications of these V-gene sequences. The regions of maximum variability occur in the three hypervariable regions which have been charted in the variability plots of Kabat¹. Each of the hypervariable regions occurs within or around the antigen binding site² and so each probably plays a part in the variability of antibodies. Recent studies of mouse L-chain genes have shown that the V segments in germ-line DNA are separate from the C-genes in the genome^{3–5}, and furthermore, that several J segments occur (equivalent to amino acid residues 96–108 (refs 6–8)) separated from the C_κ segment by a large intervening sequence. The integration of the immunoglobulin gene specifically involves fusion of one of a number of V segments with one of the J segments^{6–8} while the large intervening sequence is maintained^{4,7,8}. In the case of κ L-chain genes, residue 96 seems to be highly variable as a direct consequence of V–J joining^{7,8}. This residue, and thus, this event explains at least some of the variability of the third hypervariable region in κ-chains. In addition, it has been argued that variability of the first two κ-chain hypervariable regions and the remainder of the third hypervariable region are entirely encoded by germ-line genes⁹ whereas the λ L-chain seems to undergo some V-segment somatic variation¹⁰. There are no data for H-chain V-genes, and a first approach is to assess the variability of germ-line DNA by hybridisation experiments designed to detect the whole V_H-gene complement of a subgroup. Ideal probes for

such experiments are plasmid clones containing complementary DNA (cDNA) copies of immunoglobulin mRNA¹¹ which can be used in the filter hybridisation procedure of Southern¹². Thus, hybridisations can be carried out with suitable V_H probes and the number of hybridisation bands obtained can be equated with V-genes complementary to this probe. To assess the relevance of this number to the V-genes of a protein subgroup (that is, the group of V-region sequences bearing similar or identical framework (non-hypervariable region) amino acid residues), it is necessary to know the extent of V-gene cross-hybridisation in filter hybridisation experiments. In the present experiments, we have used a cloned cDNA probe of a mouse V_H sequence belonging to the V_HIII subgroup to determine the number of germ-line V_H-genes in this subgroup of mouse and human DNA. In the conditions used, we find that the probe hybridises to about 10 V_H-genes in mouse DNA and about 20 V_H-genes in human DNA. These numbers probably represent the size of the V_HIII subgroup in the respective DNAs, for by studying the cross-hybridisation of the κ V-gene sequences from MPC11 and MOPC21 mouse myelomas, we conclude that all members of the V_HIII subgroup should cross-hybridise.

Extent of V-gene cross-reaction in filter hybridisation

A characterisation of V-gene cross-hybridisation can be facilitated by comparing two V-genes of similar origin which differ in both the framework and hypervariable residues. The mouse κ-chains MPC11 and MOPC21 are suitable for such an analysis because protein sequencing of both chains^{13,14} shows that the V segments differ by approximately 20% of residues (excluding the J region and an additional region of 12 amino acids present

* Present address: Free University Brussels, Institute of Molecular Biology, 65 Paardenstraat, B-1640 St.-Genesius-Rode, Belgium.

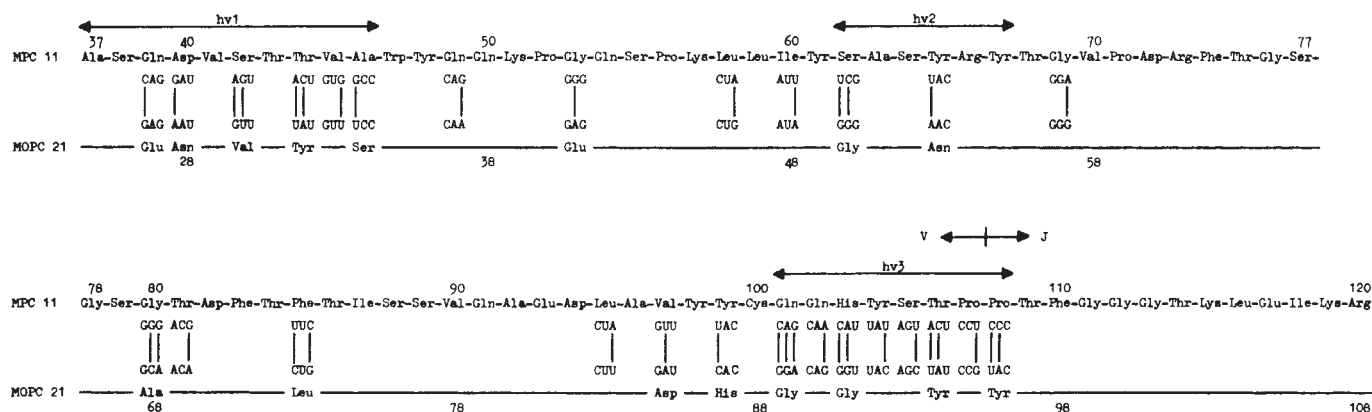


Fig. 1 Comparison of MPC11 and MOPC21 V_{κ} sequences. The nucleotide sequence of MPC11 V_{κ} shown was derived by the chemical degradation procedure¹⁶ using the 380-base pair *Hind*II fragment derived from $\mu\kappa(11)^{24}$ as a source of V-region material. This fragment was labelled with 32 P (using polynucleotide kinase) redigested with *Hinf* and *Hpa*II. The resulting fragments were isolated from acrylamide gels and subsequently subjected to degradation and gel analysis²⁷. Protein sequences were taken from Svasti and Milstein¹³ and Smith¹⁴. The MPC11 κ -chain is 12 amino acids longer than MOPC21 at the N terminus of the mature protein; each peptide chain is numbered here accordingly. Therefore, the J segment, for example, extends from amino acid 108 to 120 in MPC11 and 96 to 108 in MOPC21. Where a codon differs between MOPC21 and MPC11, the full codon is given and the bases which are different within the codon are indicated by a vertical line. All other codons are identical but some ambiguities in the nucleotide data remain: we have not been able to assign convincingly bases to the third residues of MPC11 codons 70 (Val) or 75 (Thr), nor to the third residue of MOPC21 codons 41 (Gln), 47 (Leu) or 77 (Ser). hv, Hypervariable region.

at the amino terminus of the MPC11 κ -chain¹⁴). We have compared the nucleotide sequences of these V-regions and the results of this comparison are shown in Fig. 1 (note that the additional N-terminal residues of MPC11 necessitate the use of an alternative numbering system in MPC11 and MOPC21 in this figure). The sequence data contained in Fig. 1 extend from the codon corresponding to amino acid 25 of MOPC21 (residue 37 on MPC11 numbering) through the V-J junction (residues 95–96 on MOPC21 numbering) to the end of the J segment (residue 108). The nucleotide sequences were obtained by primed reverse transcriptase sequencing of MOPC21 L-chain mRNA¹⁵ or the chemical degradation sequencing methodology¹⁶ using *Hind*II restriction fragments prepared from a cDNA clone of MPC11 κ -chain [$\mu\kappa(11)^{24}$] (see Fig. 2) containing the MPC11 V-region¹⁷. The percentage of amino acids different in the V segment of the protein sequence compared in Fig. 1 is ~23% whereas the nucleotides differ by ~18%. This proportion of difference is representative for the V-regions as the N termini of the chains (not compared in Fig. 1) also differ by about 20% of their amino acids^{13,14}.

The distribution of nucleotide differences within the two proteins is interesting from an evolutionary point of view. The germ-line progenitors of these V-genes would seem to be derived from separate V-genes which presumably had a common evolutionary origin before V-gene duplication. The framework residues of the two genes have subsequently drifted far less than the hypervariable regions and indeed, much of this drift occurs at the nucleotide level (as manifested by silent base change) but not at the amino acid level. The hypervariable regions, on the other hand, exhibit considerable variation in both amino acid changes and silent nucleotide changes. This pattern of difference between related V-genes suggests that evolutionary pressure is exerted to change the antibody combining site (the hypervariable regions) and favours the conclusion that antibody variability in mouse κ -chain results extensively from evolutionary pressures by the creation of stable V-genes rather than by extensive somatic mutation at the level of individual antibody-producing cells¹¹.

The nucleotide comparison shown in Fig. 1 indicates the type of pattern of difference we can expect between related but different V-genes. We have, therefore, considered whether such V-genes would cross-hybridise in filter hybridisation experiments using different areas of the MPC11 V-region sequence from the cDNA plasmid $\mu\kappa(11)^{24}$ (Figs 3, 4). Restriction enzyme mapping and nucleotide sequencing of $\mu\kappa(11)^{24}$ (ref. 18) and nucleotide sequence data of the C_{κ} sequence¹⁵ yield the partial restriction map of $\mu\kappa(11)^{24}$ shown in Fig. 2. The enzyme *Hind*II

cleaves the $\mu\kappa(11)^{24}$ DNA to produce, among others, a 380-base pair fragment which extends from the codon for amino acid residue 10 (using the Smith numbering¹⁴) through the whole V-region of MPC11 to the codon for amino acid 137 within the C_{κ} -region; thus, this V fragment consists of V, J and part of the C_{κ} segment. *Mbo*II cleaves this 380-base pair V fragment twice—at the codon for amino acid 96 (Smith numbering) to yield a fragment of 260 base pairs (V-region sequences only), and at the codon for amino acid 127 to give smaller fragments of about 100 base pairs (consisting of the portion of V-region containing the third hypervariable region, the J segment plus the C_{κ} segment) and about 30 base pairs (consisting of C_{κ} only).

Figure 3 shows the hybridisation of 32 P-labelled cDNA made from MOPC21 κ -chain mRNA (made with reverse transcriptase using the specific primer T₂G₃T (ref. 19)) with *Hind*II restriction fragments of $\mu\kappa(11)^{24}$, the isolated intact 380-base pair *Hind*II V fragment or this V fragment cleaved by *Mbo*II. The restriction digests (see Fig. 3 legend) were fractionated on a 2% agarose gel before binding to a cellulose nitrate filter¹² for the hybridisation reaction. *Hind*II-digested $\mu\kappa(11)^{24}$ shows two major bands of hybridisation to the heterologous MOPC21 κ cDNA (slot b), the smaller of the two components being the 380-base pair V fragment described above. The isolated V fragment (prepared by fractionation of *Hind*II digested $\mu\kappa(11)^{24}$ DNA on a polyacrylamide gel followed by elution of the 380-base pair fragment) electrophoresed in slot a, shows similar

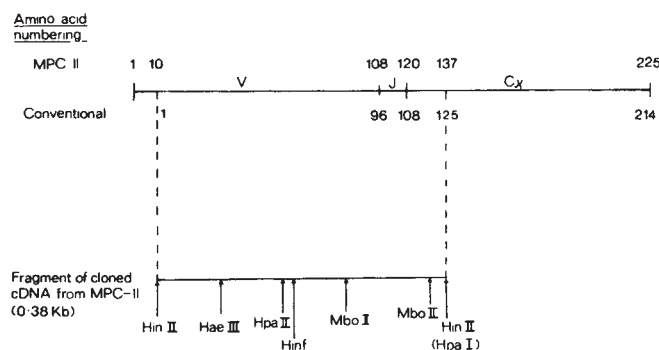


Fig. 2 Partial restriction map of the 380-base pair fragment from the MPC11 clone $\mu\kappa(11)^{24}$. The κ L-chain protein is depicted together with the partial restriction map of a *Hind*II fragment of $\mu\kappa(11)^{24}$ extending from the codon for MPC11 residue 10 to residue 137 (Smith numbering¹⁴). To relate this map to the data given in Fig. 1, the relative numbering positions of the 'conventional' (MOPC21) κ L-chain protein are written beneath the MPC11 numbering.

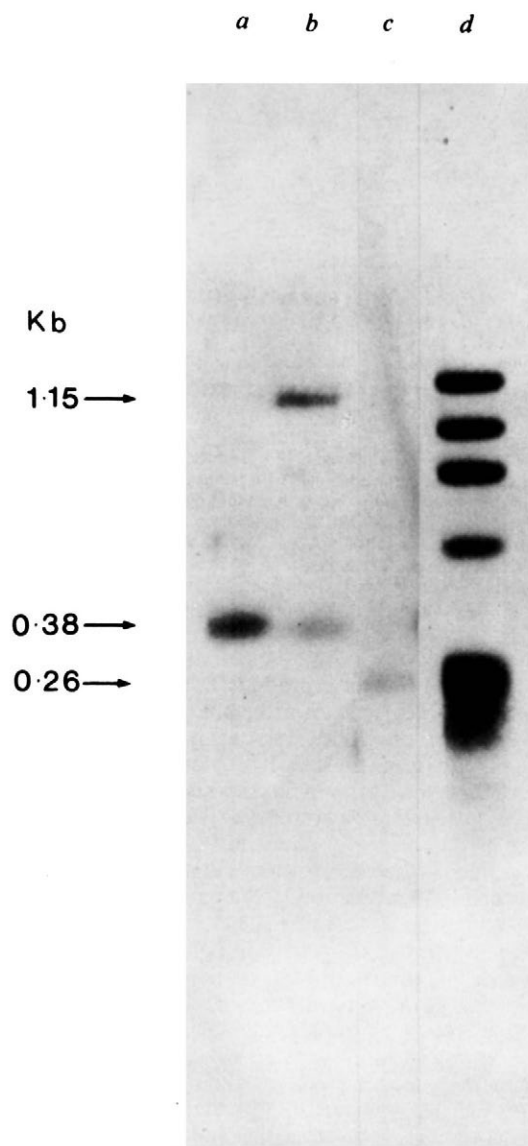


Fig. 3 Cross-hybridisation of MOPC21 κ -cDNA with *Mbo*II fragments of the isolated 380-base pair fragment from MPC11 $\rho\kappa(11)^{24}$. $\rho\kappa(11)^{24}$ DNA was digested to completion with *Hind*III, a portion fractionated on a 4% polyacrylamide gel and the 380-base pair fragment isolated¹⁸. An aliquot of this fragment was further digested with *Mbo*II. For the filter hybridisation, the digests of the various DNA samples were fractionated on a 2% agarose gel, and the DNA was denatured *in situ* and transferred to a sheet of cellulose nitrate¹² (Schleicher and Schüll BA-85). The filter was treated before hybridisation as described previously^{28,29} and hybridised for 48 h at 65 °C in 6×SSC containing 0.1% SDS, 0.2% Ficoll 400, 0.2% bovine serum albumin, 0.2% polyvinylpyrrolidone 360, 50 $\mu\text{g ml}^{-1}$ sonicated and denatured salmon sperm DNA plus 10 $\mu\text{g ml}^{-1}$ poly(rA) using 5×10^6 c.p.m. of ^{32}P -cDNA made on MOPC21 κ L-chain mRNA (by reverse transcriptase and T₂G₃T primer, specific activity of cDNA = 10^7 c.p.m. per μg (ref. 19)). After hybridisation, the filter was washed at 65 °C (30 min per wash) once with 6×SSC containing all the ingredients of the hybridisation except ^{32}P -cDNA, followed by five washes of 1×SSC and 0.1% SDS, after which the filter was rinsed in 2×SSC, dried and autoradiography carried out overnight using prefogged X-ray film³⁰. DNA samples fractionated on agarose and hybridised: a, isolated 380-base pair *Hind*III fragment of $\rho\kappa(11)^{24}$ (4 ng); b, complete *Hind*III digest of $\rho\kappa(11)^{24}$ (10 ng); c, *Mbo*II digest of the isolated 380-base pair *Hind*III fragment (10 ng); d, *Hae*III-digested ϕX174 DNA, labelled with ^{32}P before electrophoresis to serve as internal size standards in the DNA fractionation and transfer. kb, kilobase.

hybridisation to this region. On the other hand, the 380-base pair V fragment displayed a new, more rapidly migrating hybridisation component (260 base pairs) when cleaved with *Mbo*II. This component corresponds to the V segment of MPC11 (from which J and C_κ segments are absent) cross-hybridising to MOPC21 V-region cDNA. (The 120-base pair fragment also resulting from *Mbo*II cleavage of the 380-base pair V fragments

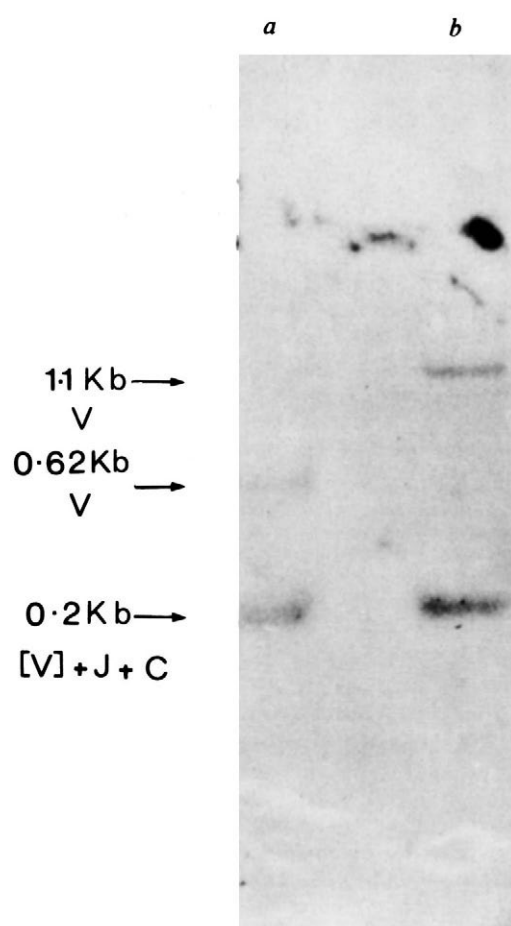


Fig. 4 Cross-hybridisation of MOPC21 κ -cDNA with *Hinf*I and *Hpa*II digests of the isolated 3.1-kilobase *Hpa*I fragment from $\rho\kappa(11)^{24}$. $\rho\kappa(11)^{24}$ DNA was cleaved by *Hpa*I, fractionated on 0.8% agarose and the 3.1-kilobase fragment eluted from the gel⁴. Aliquots (20 ng) of this fragment (containing V, J and part of the C) were redigested with either *Hinf*I (slot a) or *Hpa*II (slot b) and these digests were fractionated on a 2% agarose gel, transferred to a cellulose nitrate filter¹² and hybridised to 3×10^6 c.p.m. MOPC21 ^{32}P - κ L-chain cDNA as described in Fig. 3 legend. Sizes shown in kilobases (kb) were determined relative to unlabelled ϕX174 DNA (digested with *Hae*III) which was visualised after staining with ethidium bromide.

(which includes J and C_κ segments) does not show in the hybridisation because it was too small to bind effectively to the cellulose nitrate filter.)

Thus, the two V segments (MOPC21 and MPC11) do cross-hybridise in the conditions used for these experiments and we can, therefore, reasonably conclude that V-genes up to about 18% different at the nucleotide level will cross-hybridise. Interestingly, the MPC11 and MOPC21 V sequences from the 5' side of the *Mbo*II cleavage site (between Ala-Val position 97 (ref. 14)) contain only two substantial regions of perfect homology, one 29 bases long (Thr-86 to Leu-95) and the other 31 bases long (Val-70 to Gly-80). These areas represent the best possibilities for cross-hybridisation of the V-genes. The remainder of the sequences are broken up by base differences of various complexity. In view of this, we examined further the V-gene cross-hybridisation with segments of the V-region which exclude the homologous stretches of 29 and 31 nucleotides.

Restriction mapping of $\rho\kappa(11)^{24}$ showed the presence of sites for the restriction enzyme *Hinf*I between the codons for Gly-Val at positions 69–70 and *Hpa*II between Tyr and Arg at positions 65–66 (Fig. 2). Thus, by cleaving $\rho\kappa(11)^{24}$ with either of these enzymes it is possible to generate a V segment suitable for cross-hybridisation studies which excludes the relatively long stretches of homologous sequence but includes the second hypervariable region. To simplify the analysis of the cross-hybridisation of these V segments, we prepared a subfragment of $\rho\kappa(11)^{24}$ with which to carry out the intra-V-segment

cleavage. The nucleotide sequence of the C_{κ} -region showed a cleavage site for *HpaI* at the codon for amino acid 125 (ref. 15). *HpaI* cleaves $\rho\kappa(11)^{24}$ at this site and at a site in the pMB9 vector to yield two large fragments of 3.7 and 3.1 kilobases. The 3.1-kilobase fragment contains the V segment together with the J and the 5' end of the C_{κ} . When this isolated 3.1-kilobase fragment is digested with *Hinf* or *HpaII*, the V segment is split into two fragments of 0.61 and 0.2 kilobases with *Hinf* cleavage and 1.1 and 0.2 kilobases with *HpaII* cleavage (unpublished data). In both cases, the 0.2-kilobase fragment is the *HpaI*-*Hinf* or *HpaI*-*HpaII* fragment containing the 5' end of the C_{κ} -region, the J plus the 3' end of the V segment (Fig. 2). The larger of the two fragments then consists of the 5' end of the V-region starting at around the second hypervariable region.

Figure 4 shows the hybridisation of purified MOPC21 L-chain cDNA with the *Hinf* or *HpaII* digests of the 3.1-kilobase *HpaI* fragment of the MPC11 clone $\rho\kappa(11)^{24}$. In both cases the large and small V segment-containing fragments hybridised to the MOPC21 cDNA. When the *HpaI* fragment was cleaved by *Hinf* we observed that both the 0.2- and 0.62-kilobase fragments hybridised the MOPC21 cDNA (slot a) and similarly the 0.2- and 1.1-kilobase fragments of the *HpaII* digest hybridised to the MOPC21 cDNA (slot b). Thus, again we observe that the fragment of the MPC11 clone which contains the V segment hybridised to the heterologous MOPC21 cDNA. In addition, in this case, we detect hybridisation between the V segments where no more than 18 uninterrupted base pairs can form. (For example, in the area of nucleotide sequence shown in Fig. 1, 15 homologous base pairs can form between codons for Gln-54 and Leu-58 whereas in the region where only protein sequence is known a possible 18-base pair region could exist between Ser-26 and Val-31.) The results of these cross-hybridisation experiments show that V_{κ} -genes with only short areas of homologous base sequence, interrupted extensively by mismatches, do cross-hybridise in the filter hybridisation experiments. It seems reasonable to assume that this situation will also occur with all related V-genes as the patterns of variability in general follow that shown by MPC11 and MOPC21 V_{κ} -regions. These cross-hybridisation experiments therefore set a criterion for the extent of V-gene detection and we can say that genes differing by at least 18% of their nucleotides can be detected in this type of experiment.

It must be considered, however, that the detection of cross-hybridising V segments when using cloned DNA, need not necessarily be related to such cross-hybridisations in genomic DNA. The experiments described in Figs 3 and 4 used 0.5 ng, 2 ng and 4 ng of specific hybridising sequence compared with the expected quantity of about 0.02 ng of a particular V-gene in a genomic hybridisation. This means that we are using between 25 and 200 times as much cloned V segment in these controls as

used in the genomic blots described below. This could lead to underestimates of V-gene numbers due to lower sensitivity of the latter. However, it is also true that the sensitivity of detection of the small DNA fragments described in Figs 3 and 4 is considerably weakened (compared with that of large fragments) by the relative inefficiency of binding small fragments to the cellulose nitrate filters. On balance, the relatively high quantities of the small cloned fragments are likely to mimic the situation in the genomic hybridisations.

The number of V_H -genes in mouse and human DNA

When assaying for V-genes in genomic DNA by filter hybridisation, the genes detected must represent either a group of closely related V-genes separated from the next set by at least 18% nucleotide difference, or V-genes gradually diverging in sequence from one another up to the extent of or greater than about 18% difference. Whichever of these situations exist, the results obtained for the DNA of a given species allows an assessment of the size of the inherited V-gene pool and in turn the potential requirement for somatic variation of the V-genes. We have attempted to derive information relating to V_H -gene complexity in mouse and human DNA using a mouse cDNA plasmid probe containing V_H -gene sequences. A cDNA plasmid ($\rho\mu/107$) has been derived from μ H-chain mRNA (of the hybrid cell line Sp1/HL (ref. 20)) and this plasmid has been shown to contain the entire V_H and C_{μ} sequences²¹. The nucleotide sequence of the 5' end of the V_H segment of $\rho\mu/107$ has been derived and is shown in Fig. 5 (extending from amino acid codons 1 to 38). This V_H -region (an anti-sheep red blood cell antibody) can be assigned to the mouse V_{HIII} subgroup, from which it differs by two amino acids in the first framework region from the basic V_{HIII} sequence¹ and a further four residues in the first hypervariable region (16% difference).

As shown in Fig. 5, the other two well-characterised mouse V_H subgroups [V_{HI} and V_{HV} (ref. 1)] differ extensively from the anti-sheep red blood cell Sp1/HL V_H sequence. V_{HI} differs by about 52% and V_{HV} by about 42% from mouse V_{HIII} . By considering the cross-hybridisation capabilities of V-genes analysed in the previous section, we can see that the V_H sequence in $\rho\mu/107$ would be expected to cross-hybridise to all V_{HIII} members because related V_{HIII} genes are likely to display the same type of variability at the nucleotide level as seen for the related V_{κ} sequences¹. The V_H -genes would be able, therefore, to cross-hybridise by virtue of relatively small regions of uninterrupted base pairing. On the other hand, comparison of the $\rho\mu/107$ V-region with the other subgroup sequences indicates that cross-hybridisation is unlikely because, on the basis of the sequences shown we can only find one region of potentially 15

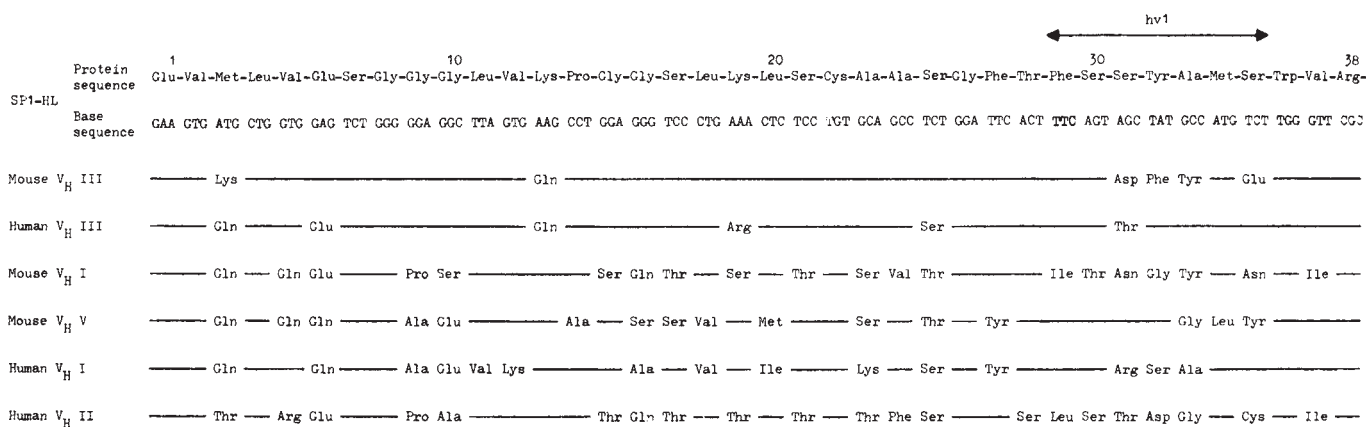


Fig. 5 Sequence of 5' terminus of the Sp1/HL heavy-chain anti-SRBC specificity and comparison with mouse and human V_H subgroups. The nucleotide sequence shown for Sp1/HL was derived by chemical degradation procedures of Maxam and Gilbert¹⁶ using an *EcoRI*-*HpaII* restriction fragment from $\rho\mu/107$ (ref. 21). The derived protein sequence is compared with the basic sequences of characterised mouse and human V_H subgroups¹.

homologous base pairs (in V_HV). We have also compared the Sp1/HL sequence with the human V_H subgroup basic sequences (Fig. 5). This comparison revealed that the Sp1/HL sequence is also very similar to human V_{HIII} (about 16% difference between residues 1 and 38) and, further, is more similar to the Sp1/HL sequence than the basic V_{HIII} of mouse in the area designated as the first hypervariable region. Thus, we should expect to achieve intra-subgroup cross-hybridisation between the mouse V_{HIII} probe and the human V_{HIII} genes, whereas the other human V_H subgroups differ to the extent that little cross-hybridisation would be expected (for example, there is about 39% difference between the Sp1/HL V sequence and human V_{HI} and 55% between Sp1/HL V_H and human V_{HII}).

Therefore, by carrying out filter hybridisation of mouse and human DNA with the $p\mu/107$ probe, we can estimate the V_H -gene complement for the V_{HIII} subgroup in these species. Figure 6 shows the result of filter hybridisation between ^{32}P -labelled $p\mu/107$ and fractionations of *EcoRI*-digested BALB/c mouse liver DNA (slot *a*) and *EcoRI*-digested human placental DNA (slot *b*). The hybridisation of ^{32}P -labelled $p\mu/107$ to mouse liver DNA yielded around 10 bands of hybridisation ranging from about 12.5 kilobases to about 2.7 kilobases. We have identified the largest hybridising band (12.5 kilobases) as containing the C_μ -gene so that the remaining bands represent V_H -gene hybridisation (unpublished results). In addition, it seems likely that each band represents only one hybridisation band rather than co-electrophoreses of fragments of similar size, as digests of mouse DNA with different enzymes yielded a similar number of hybridising restriction fragments. Previous studies of mouse V_κ -genes has indicated that, in general, V_κ -genes are at least 10 kilobases apart²², and our own studies on isolated V_H -genes from human DNA indicates that the average spacing of V_H -genes is greater than 10 kilobases (unpublished data). It is, therefore, a reasonable conclusion that each band contains a single V_H -gene. Thus, there are only about 10 V_H sequences in the mouse germ-line DNA which fall within the detection range of 18% nucleotide difference, and the mouse V_{HIII} subgroup therefore seems to contain only about 10 V_H -genes. Protein sequence data indicate that only four or five V_H subgroups occur in mouse¹ and assuming that they are all of a similar size, we expect about 40–50 V_H -genes in total. This number compares with the projected V_κ -gene pool size of 300 predicted from a similar type of hybridisation experiment¹¹.

Figure 6 also shows the result of hybridising the $p\mu/107$ probe to *EcoRI*-digested human placental DNA. Here we observed a more complex pattern than with mouse DNA, the hybridisation profile showing around 20 hybridisation bands ranging from about 17 kilobases to about 1 kilobase. Two of these bands hybridise to a C_μ probe¹⁸, so we are observing about 18 hybridising V_H -genes (equating each hybridising component with individual V_H -genes by analogous arguments to those for the mouse). Therefore, as we have argued that we can detect all or most V_H -genes in the human V_{HIII} subgroup, this subgroup would seem to consist of around 20 V -genes. The V_H -gene pool of human DNA would therefore seem to consist of around 80 V -genes as protein data show that there are four human V_H subgroups¹. The projected V_H -gene pool sizes calculated from the hybridisation data assume that all the protein subgroups have been discovered and that each subgroup is of approximately equal size. Discrepancies in these two parameters will, of course, introduce errors into the final value. However, study of available protein data suggests that the V_{HIII} subgroup may be a major representative of the germ-line component of V_H -genes¹.

Origin of variability of heavy-chain V-genes

The hybridisation data presented here suggest that the V_{HIII} subgroup of mouse and human DNA contains a low number of genes and that the V_H -gene pool is of the order of 40–80 genes in both species. The C_μ -gene, $C\gamma 2b$ and $C\gamma 2a$ (refs 18, 23, 26) and $C\gamma 1$ (ref. 24) genes, on the other hand, seem to represent unique components in the genome. Therefore, two clear conclusions

can be drawn. First, the V_H -gene pool must lie in the germ-line DNA separate from the C_H -gene array as in the case of the L-chains^{3–5}. Second, in order to express a V_H -gene in combination initially with the C_μ -gene, a translocation recombination event must occur to form the active H-chain gene transcription unit.

Antibody variability results from a combination of H- and L-chains and therefore the variability ultimately attained is dictated by the germ-line complement of V_H and V_L . The mouse and human V_{HIII} subgroups seem to be approximately the same size although the human subgroup is slightly larger than the corresponding group of mouse. The combination of V_H and V_κ segments in mouse could potentially generate 1.5×10^4 antibody specificities. We have no data on the V_κ complement of human DNA, but it seems likely from the available protein data that the number of κ V-genes will be smaller in human than mouse. The combination of the V_H -gene pool potential with V_κ -genes in human would, therefore, seem to be lower than the correspond-

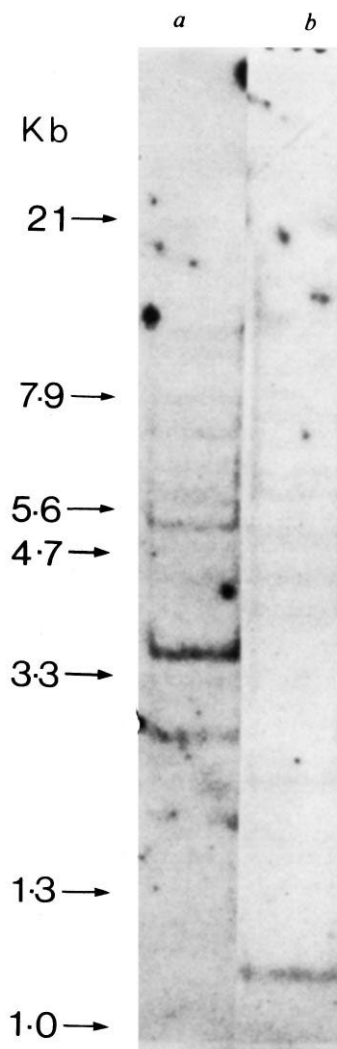


Fig. 6 Filter hybridisation patterns of the mouse heavy-chain cDNA probe $p\mu/107$ with *EcoRI*-digested BALB/c mouse liver DNA (slot *a*) and human placental DNA (slot *b*). 15 μ g of mouse liver or human placental DNA were completely digested with *EcoRI* and fractionated on an 0.8% agarose gel alongside a mixture of *EcoRI* digest of λ phage DNA (size range 21–3.3 kilobases) and *HaeIII*-digested Φ X174 (fragments visible were 1.3 and 1.0 kilobases). The positions of these markers were visualised under UV light after staining with ethidium bromide and the positions are indicated on the left hand side. After staining, the DNA was transferred to a cellulose nitrate filter¹² and the filter hybridised (as described in Fig. 3 legend) for 48 h with 25 ng ml⁻¹ ^{32}P -labelled $p\mu/107$ (labelled by nick-translation^{31,32} to a specific activity of 5×10^7 c.p.m. per μ g), followed by post-hybridisation washing and autoradiography as Fig. 3.

ing combination in mouse. This would imply that somatic variation of V segments (both V_H and V_κ) has a more important role in the generation of diversity in V segments in the human species than in the mouse. In addition, the apparently lower number of V_H -genes in mouse germ line compared with the corresponding V_κ -gene number, indicates that somatic variation of V_H -genes may be more significant than in V_κ -genes of mouse.

It is possible, however, that, particularly in mouse, the $H \times L$ combination can generate all required antibodies without significant somatic variation in the V segments. It has been postulated that the single residue 96, which is involved in the translocation joining of V_κ and J_κ , may be highly variable as a result of this joining^{7,8}, so that a somatically produced hyper-variation is believed to occur in κ -chains of mouse. It is not known whether J_H segments exist in mouse or human, although some evidence for their existence has been proposed²⁵. Therefore, the effective number of V_H -regions randomly attainable

will be the product of V_H and putative J_H segments, assuming the V and J can fuse randomly. In this respect, it is interesting that the third hypervariable (hv_3) region may be largely contained in the J_H -region, unlike the analogous region of κ -chains which does not contain much of hv_3 (refs 7, 8). Therefore, the overall number of different H-chain V-regions which can be produced will depend very much on the number of J_H segments available, as the fusion of V_H with J_H would combine the first two hv_3 regions with the different hv_3 regions within the J_H . In addition, the sequence of Sp1/HL V-region shows that it differs significantly from the basic V_H III sequence in both framework residues and the first hypervariable region, implying that some somatic variation may be operating on these sequences.

We thank Dr R. Perry for supplying the MPC11 clone $\mu\kappa(11)^{24}$. G.M. was the recipient of fellowships from the Royal Society and the National Fonds voor Wetenschappelyk Onderzoek, Belgium.

Received 16 November 1979; accepted 23 January 1980.

- Kabat, E. A., Wu, T. T. & Bilofsky, H. in *Sequences of Ig Chains* (NIH Publications, 1979).
- Givol, D. *Int. Rev. Biochem.* **11b**, 70–131 (1979).
- Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976).
- Rabbitts, T. H. & Forster, A. *Cell* **13**, 319–327 (1978).
- Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3881–3885 (1978).
- Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
- Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
- Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454 (1979).
- Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370–375 (1979).
- Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133–1144 (1978).
- Seidman, J. G. & Leder, P. *Nature* **276**, 790–795 (1978).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Svasti, J. & Milstein, C. *Biochem. J.* **128**, 427–444 (1972).
- Smith, G. P. *Biochem. J.* **171**, 337–347 (1978).
- Hamlyn, P. H., Brownlee, G. G., Cheng, C.-C., Gait, M. J. & Milstein, C. *Cell* **15**, 1067–1075 (1978).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
- Schibler, U., Marcu, K. B. & Perry, R. P. *Cell* **15**, 1495–1509 (1979).
- Rabbitts, T. H., Hamlyn, P. H., Matthysens, G. & Roe, B. A. in *Symp. Xlth int. Cong. Biochem.* (ed. Kaplan, J. G.) (in the press).
- Rabbitts, T. H., Forster, A., Smith, M. & Gilham, S. *Eur. J. Immun.* **7**, 43–48 (1977).
- Köhler, G. & Milstein, C. *Eur. J. Immun.* **6**, 511–519 (1976).
- Matthysens, G. & Rabbitts, T. H. *Nucleic Acids Res.* (in the press).
- Lenhard-Schuller, R., Hohn, B., Brack, C., Hiram, M. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4709–4713 (1978).
- Kataoka, T., Yamawaki-Kataoka, Y., Yamagishi, H. & Honjo, T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4240–4244 (1979).
- Sakano, H. *et al. Nature* **277**, 627–633 (1979).
- Rao, N., Rudikoff, S., Krutzsch, H. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2890–2894 (1979).
- Rabbitts, T. H., Forster, A., Dunnick, W. & Bentley, O. L. *Nature* **283**, 351–356 (1980).
- Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107–110 (1978).
- Denhardt, D. T. *Biochem. biophys. Res. Commun.* **23**, 641–646 (1966).
- Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097–1108 (1977).
- Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82**, 314–316 (1977).
- Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).

LETTERS

Spectrum of the Venus day sky

V. I. Moroz, Yu. M. Golovin, A. P. Ekonomov,
B. E. Moshkin, N. A. Parfent'ev & N. F. San'ko

Space Research Institute, USSR Academy of Sciences, Moscow, USSR

The spectrum of solar radiation reflected by Venus' atmosphere has long been studied by ground-based telescopes: these studies were a major source of information on the composition of atmosphere gases and also about the upper part of the cloud layer. In December 1978 the Soviet descenders Venera 11 and Venera 12 measured the spectrum of scattered solar radiation within the atmosphere with a resolution sufficient to identify the absorption bands of atmosphere gases. We report here the first observations of the spectra of the Venus day sky.

Both vehicles landed in the equatorial region of the planet; the solar zenith angle was about 20°. The coordinates and other details of the landing area are given elsewhere¹. The results mentioned below were obtained using scanning spectrophotometers on-board the descenders. Each instrument had two measuring channels. Channel 1 recorded the spectrum of scattered solar radiation in the range 0.43–1.17 μm . The analysing element was a circular variable interference filter with a resolution of $\lambda/\Delta\lambda \approx 30$; the radiation was received from the zenith. During the station descent about 300 spectra were obtained between 62 km and the surface. Channel 2 measured the distribution of the radiation intensity in the vertical plane in four spectra regions: 0.45–0.55, 0.56–0.68, 0.9–1.14, and 1.15–1.55 μm separated by glass filters. The field of view of both channels was 20° at the 0.5 level. The instrument and its

operating conditions are described in more detail elsewhere². Figure 1 shows the spectra recorded by the Venera 11 spectrophotometer; also shown, for comparison, is the solar radiation spectrum.

The spectra at 62, 56 and 51 km were obtained inside the cloud layer. The intensities of these spectra diminish rapidly as the altitude decreases. After the vehicle emerges from the clouds (at ~49 km) the intensity varies very weakly down to 38 km and then the rapid decrease of intensity is observed as the atmosphere optical depth increases. The absorption bands of CO_2 and H_2O are seen in the spectra. The spectra measured by the two vehicles are in good agreement.

To estimate the integral flux F on the surface, the integration was made over the wavelength range 0.43–1.17 μm and over the entire upper hemisphere. It was found that $F = 72 \text{ W m}^{-2}$, which is ~3% of the solar radiation flux outside the atmosphere. The surface illuminance is about 9,000 lx, which is in good agreement with the results from Venera 9 and Venera 10 (ref. 3). According to the spectrum obtained, the colour of the surface is orange.

Figure 2 gives the height dependence of radiation intensities from the zenith for $\lambda = 0.45, 0.5, 0.635, 0.72$ and $1.0 \mu\text{m}$. The part of the curve for $\lambda = 0.5 \mu\text{m}$ is shown on a larger scale, together with experimental points. Examination of the curves shows three sublayers with different extinction coefficients: a 49–52 km sublayer is the most dense optically, 52–58 km is the most transparent and 58–62 km forms a layer with an intermediate optical density. Nearly the same stratification was found in the optical experiments on Pioneer Venus^{4–6}.

Estimates of the multiple light scattering were used to analyse the data. In one of the analyses the atmosphere was divided into 20 layers, intensities were calculated by the adding method⁷. The reflection and transmission of each layer were determined

by a two-flux approximation. The profiles of intensity, close to those measured in the continuous spectrum, were estimated using the variations layer model parameters. The following model was constructed from these results. (1) The main cloud layer is at 49–68 km, its full optical depth is $25 < \tau < 35$ in a spectral region of $0.63 < \lambda < 1.00 \mu\text{m}$; (2) the optical depth of the lower most dense part of the main cloud layer (49–52 km) is about 12; (3) the single scattering cloud particles albedo is $0.998 < a < 1.000$; (4) at 49–52 km there is very weak aerosol haze with an optical depth of about 0.7 below the main cloud layer. Optical depth has been estimated using the assumption that the scattering is anisotropic and Henyey–Greenstein phase function parameter is $g = 0.7$ (this corresponds to size of particles $\sim 1 \mu\text{m}$).

The above mentioned characteristics of the cloud layer, together with the data on the Rayleigh scattering and the absorption coefficients of H_2O and CO_2 , were used to calculate synthetic spectra in the region of the absorption bands. Comparison of synthetic and measured spectra yields a preliminary conclusion that the mixing ratio is $[\text{H}_2\text{O}]/[\text{CO}_2] \approx 2 \times 10^{-5}$ near the surface but that it gradually increases and in the clouds it is $[\text{H}_2\text{O}]/[\text{CO}_2] \approx 2 \times 10^{-4}$. The last evaluation is close to the results obtained on Venera 9 and Venera 10 (ref. 8). Both estimates strongly contradict the published results of measurements of the gas chromatograph on Pioneer Venus⁹, which give the abundance of H_2O from 0.1 to 0.5%. The gas chromatograph on Venera 12 gave the upper limit of $\text{H}_2\text{O} < 10^{-4}$ for heights under 42 km that agrees with our estimates. The decrease of H_2O mixing ratio between the clouds and the surface may be associated with some processes of chemical binding of H_2O in rocks¹¹.

The depression in the blue part of the measured spectrum (heights under 30 km, Fig. 1) can be only partially explained by the Rayleigh scattering. Sulphur in the gaseous phase (S_2) can absorb in this region and in this case the mixing ratio of $[\text{S}_2]/[\text{CO}_2]$ is $\sim 10^{-8}$. In any case this value can be considered as an approximate upper limit. There may possibly be a contribution from absorption of other molecules here, such as Cl_2 , Br_2 and NO_2 . The corresponding upper limits are approximately 10^{-8} , 10^{-10} and 5×10^{-10} . The sensitivity of the optical spectroscopy for some substances (even for low resolutions) in the Venus low atmosphere turns out to be higher than for the mass

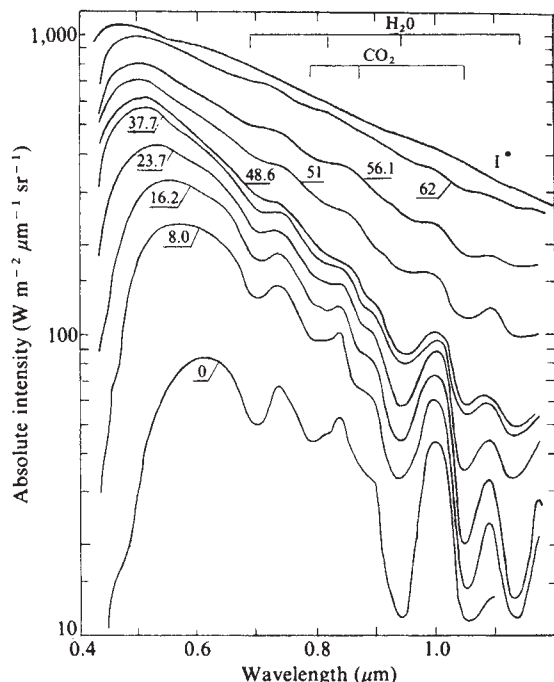


Fig. 1 Samples of spectra of scattered solar radiation obtained by Venera 11. Numbers near the curves are heights above the surface.

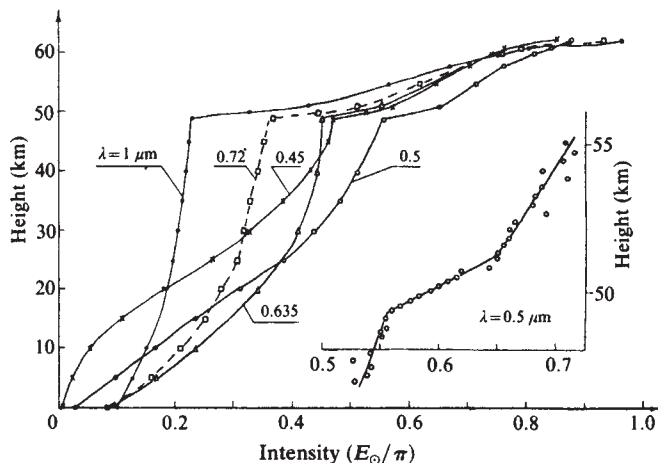


Fig. 2 Zenith intensity of the radiation as a height function for some wavelengths. E_0 is the solar illuminance at the upper boundary of the Venus atmosphere. Numbers near the curves are the wavelengths.

spectrometer and the gas chromatograph. There is an indication that the concentration of the substance absorbing in the blue might decrease near the surface. In the altitude range 0–7 km the height dependence of intensity (in the blue) can be explained by assuming that there is only the Rayleigh scattering in this region of the atmosphere; the absorption appears only at higher altitudes. This absorption practically disappears again above 30 km. There is also some blue absorption in the top part of cloud. This cloud absorption, taking the assumption¹² into account, is caused by condensed sulphur.

More detailed reports on the above analysis of the spectra obtained on Venera 11 and Venera 12 probes are given elsewhere^{2,13,14}.

Received 1 November; accepted 7 December 1979.

1. Pis'ma Astr. Zh. 5, 333 (1979).
2. Ekonomov, A. P. et al. *Kosm. Issled.* 17, 714 (1979).
3. Moshkin, B. E., Ekonomov, A. P. & Golovin, Yu. M. *Kosmich. Issled.* 16, 412 (1978).
4. Ragert, B. & Blamont, J. *Science* 203, 790 (1979).
5. Tomasko, M. G. *Science* 203, 795 (1979).
6. Knollenberg, K. G. & Hunt, D. M. *Science* 203, 792 (1979).
7. Lacy, A. A. & Hansen, J. A. *J. Atmos. Sci.* 31, 118 (1974).
8. Ustinov, E. A. & Moroz, V. I. *Kosmich. Issled.* 16, 127 (1978).
9. Oyama, V. I. et al. *Science* 203, 802, 1979.
10. Gelman, B. G. et al. *Pis'ma Astr. Zh.* 5, 217 (1979).
11. Khodakovskij, I. L. et al. *Geokhimiya* No. 12 1821 (1978).
12. Young, A. T. *Icarus* 32, 1 (1977).
13. Moroz, V. I. et al. *Pis'ma Astr. Zh.* 5, 222 (1979).
14. Moroz, V. I., Parfent'ev, N. A. & San'ko, N. F. *Kosmich. Issled.* 17, 727 (1979).

Discovery of frequent lightning discharges in clouds on Venus

L. V. Ksanfomaliti

Space Research Institute, Profsoyuznaja, 84, 117810 Moscow, USSR

The analysis of the composition of the Venus atmosphere made by seven Soviet Venera-type spacecraft between 1967 and 1975 has resulted in speculation about the possible role of lightning in the formation of some minor atmospheric components similar to the Earth's atmosphere. The glowing of the Venus nightside, the ashen light, sometimes observed¹ might also be explained by lightning in the Venus atmosphere. Calculations show, however, that there must be many lightning discharges to be visible from the Earth. Nevertheless, although some speculation has arisen about the existence of lightning in the atmospheres of the other planets until recently only lightning discharges in the Earth atmosphere were known.

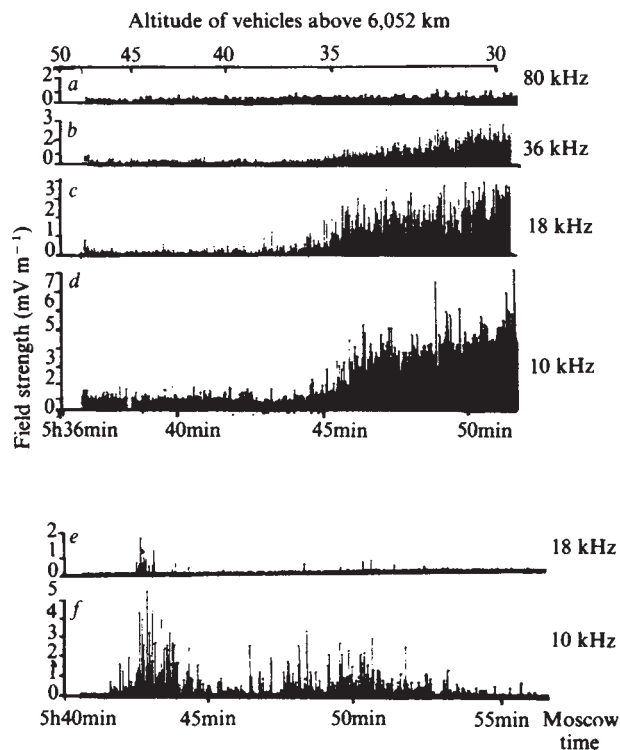


Fig. 1 Comparison of the electric thunderstorm activity of the Venus atmosphere during the Venera 12 descent (*e, f*, 21 December 1978) and the Venera 11 descent (*a-d*, 25 December 1978). The altitude scale refers to both vehicles. The signal was recorded in the bands: 1.6, 2.6, 4.6 and 15 kHz at frequencies: 10, 18, 36 and 80 kHz, respectively.

The GROZA experiment was carried out on board the Venera 11 and Venera 12 spacecraft in December 1978 to search for lightning in the Venus atmosphere². (Hara³ has developed the same idea independently, while our experiment was under preparation.) The GROZA instrument recorded impulses of the electromagnetic radiation produced by the lightning discharges. It consists of a radio receiver/spectroanalyser covering a range of 8 to 90 kHz and an impulse counter of discharge rate. The instrument had an external magnetic loop antenna operating at temperatures up to 500 °C and pressures up to 100 bar. The instrument worked both during its descent and on the planetary surface.

The following points were considered during preparation of the experiment. The intensive motions in the planetary cloud layer⁴ at heights 50–70 km were expected to contribute greatly to the accumulation of rather strong electrical charges in

clouds⁵. The large altitude of the lower boundary of the cloud layer means that the discharge to the surface would require enormous potentials. 'Cloud–cloud' discharges, well known in case of the Earth are more probable. The discharges of Earth lightning are characterised by high energies of $\sim 10^9$ J. Thus, energy is determined by the conditions in which electrical discharges occur in a gas: the energy an electron obtained during its mean free path has to exceed the gas ionisation potential (that is, of 13.8 eV for CO₂ and somewhat less for air). To satisfy this condition the electric field must reach some hundreds of kV m⁻¹. Therefore, the charges accumulating in the atmosphere reach tens of coulombs⁶. The lightning is followed by the transport of a very significant charge, equal to about several coulombs, that produces the high energies of lightnings. As for aerosols their conductivity affects discharge only slightly. Hence if there is an accumulation of charges in clouds on Venus the lightning energy must be of the same order as that on the Earth.

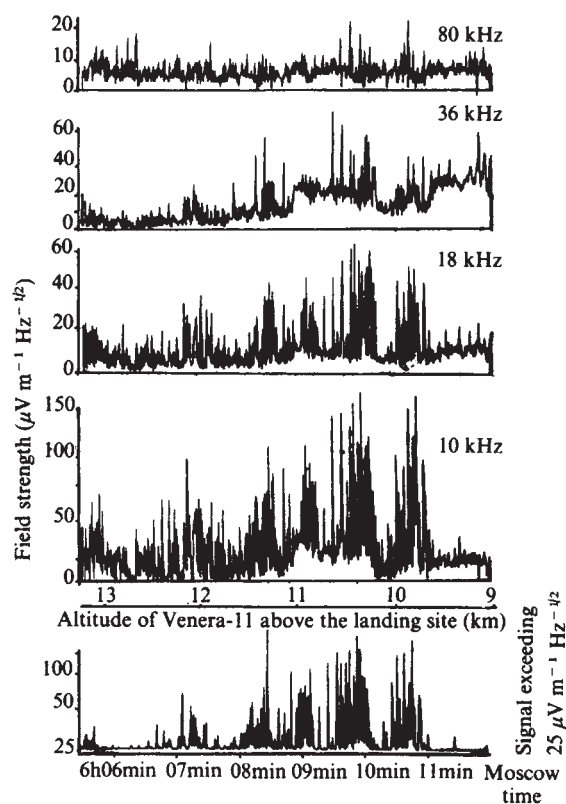


Fig. 3 Periodic sequence of the radionoise bursts recorded by the Venera 11 at heights 13–10 km. The source was the thunderstorm region with the lateral extent of ~ 150 km at $\sim 1,500$ km from the landing site.

The space vehicles had almost the same descending trajectory, and the descent operation was carried out in the same equatorial region of Venus⁷. During the descent a thousand impulses were recorded that were similar to the radiation of lightning in distant thunderstorms on Earth but with a high rate of discharge. While the Venera 11 was descending, the thunderstorm events were rather intense (on 25 December 1978) but during the descent of the Venera 12 (on 21 December 1978) there were fewer discharges. The solar activity during the latter half of December was very low and comparatively constant. The difference in the character of these events probably relates to the local nature of Venus thunderstorms. The discharge rate was 20–30 s⁻¹ and in some cases reached up to 50 s⁻¹. The comparison of the records made at heights of 30–50 km for both vehicles is shown in Fig. 1: *e, f*, Venera 12; and *a-d*, Venera 11. The field strength reaches 300 $\mu\text{V m}^{-1} \text{Hz}^{-1/2}$ at 10 kHz. The

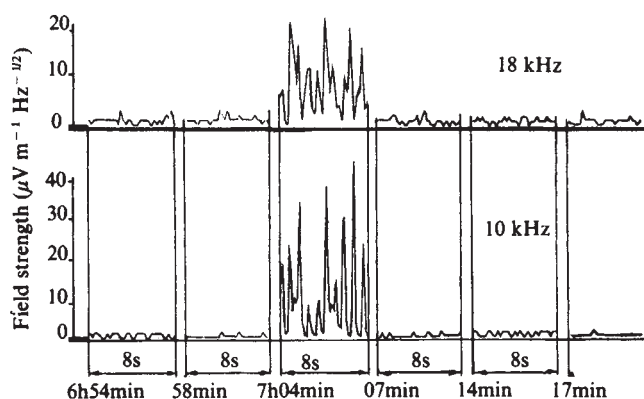


Fig. 2 The unique group of impulses recorded on the planetary surface by Venera 12, 30 min after the landing.

decrease of the strength with a frequency is close to f^{-1} on average. Due to the small time constant of the spectral bands (0.24 s) and the compressed character of the curves, each impulse record is in the form of vertical bars. Venera 11 recorded impulses continuously; five different sources (or events) can apparently be singled out. On the planetary surface the GROZA counters recorded during 110 min only one group of 160 impulses received for 8 s. Due to the sampling frequency (3 s^{-1}) the field strength at the 10 and 18 kHz bands resembled that shown in Fig. 2.

The sequence of impulses with noticeable periodicity associated with both the vehicle slow rotation and the directional pattern of the GROZA antenna turned out to be a key for determining some characteristics of venusian thunderstorms (Fig. 3). The sequence consists of six periods of great bursts with some hundreds of impulses in each. Calculations² showed that the signal was emitted by the source with small angular size of $\sim 5^\circ$. The field strength showed that the source was $\sim 1,500 \text{ km}$ away and had a lateral extent of about 150 km (it was assumed that the source strength in Venus clouds was similar to that in terrestrial clouds). Probably the source was a part of the cloud layer. There were about 25 lightnings per second in this source: this burst took place on the radio horizon and for the next few minutes the source should be behind the radio horizon—a fact which should have been confirmed by Fig. 3. However, more precise calculations using the diffraction theory showed that the radio occultation should have lasted for many minutes in spite of the rapid descent of the vehicle; therefore, it is difficult to explain the sudden cutoff of the signal at 6 h 11 min. Perhaps there are peculiarities in the wave propagation, typical of Venus, that we don't know about, as the field strength from other radio sources also decreases very quickly for 2–3 min.

From 8 km above the surface until they landed both vehicles recorded the many impulses having moderate amplitudes which decreased up to the same level as the radionoise near the surface (Fig. 4). This is the only event seen at the same altitude on the records of both 21 December and 25 December. In Fig. 4 the

group of impulses with distinct frequency properties can be singled out: from 6 h 11 min to 6 h 14 min the field strength was $50 \mu\text{V m}^{-1} \text{ Hz}^{-1/2}$ only in the range 33–38 kHz whereas at frequencies 10 and 18 kHz the field strength was only $10\text{--}20 \mu\text{V m}^{-1} \text{ Hz}^{-1/2}$ —this event also remains obscure.

From the beginning of January 1979 the Pioneer Venus orbiter began recording low-frequency impulses of electromagnetic radiation at periastris which Taylor *et al.* also identified with the lightning radiation⁸. So, these observations appear to confirm our interpretation. The high lightning rate on Venus apparently makes it possible to explain its glowing on the nightside. Calculation show that for the phase angle of 90° (when the half of the planet is illuminated) the light flow from the nightside can be 10^{-4} of the reflected sunlight if the thunderstorm activity covers 4–38% of the planetary surface depending on the efficiency of the transformation of lightning energy into light⁹.

Intensive and frequent discharges of lightning (up to 50 s^{-1} in one active region) occur in the Venus cloud layer. The origin of some gaseous components of the atmosphere can be associated with this lightning. During the Venera 11 and Venera 12 investigations the thunderstorms were of a local nature. The high lightning activity on Venus can explain the nightside glowing observed from the Earth. Lightning phenomena are, therefore, typical of the Earth and also of other planets with rather dense atmospheres and clouds where the conditions necessary for great charges to accumulate are realised.

Received 18 September; accepted 18 December 1979.

1. Kozyrev, N. A. *Izv. Krj'm. astrofiz. Obs.* **12**, 169–175 (1954).
2. Ksanfomaliti, L. V. *et al. Pis'ma Astr. Zh.* **5**, 229–236 (1979).
3. Hara, T., Preprint, University Kyoto (1976).
4. Counselman, C. C. III, Gourevitch, S. A., King, R. M., Loriot, S. B. & Brynn, R. G. *Science* **205**, 85–87 (1979).
5. Imyanitov, I. M., Chubarina, E. V. & Shvarts, Ya. M. *Elektrichestvo Oblak.* (Gidrometeoizdat, Leningrad, 1971).
6. Uman, M. A. *Lightning* (McGraw-Hill, New York, 1969).
7. *Pis'ma Astr. Zh.* **5**, 3–7 (1979).
8. Taylor, W. W. L., Scarf, F. L., Russell, C. T. & Brace, L. H. *Nature* **279**, 614–616 (1979).
9. Ksanfomaliti, L. V. *Kosmich. Issled.* **17**, 747–762 (1979).

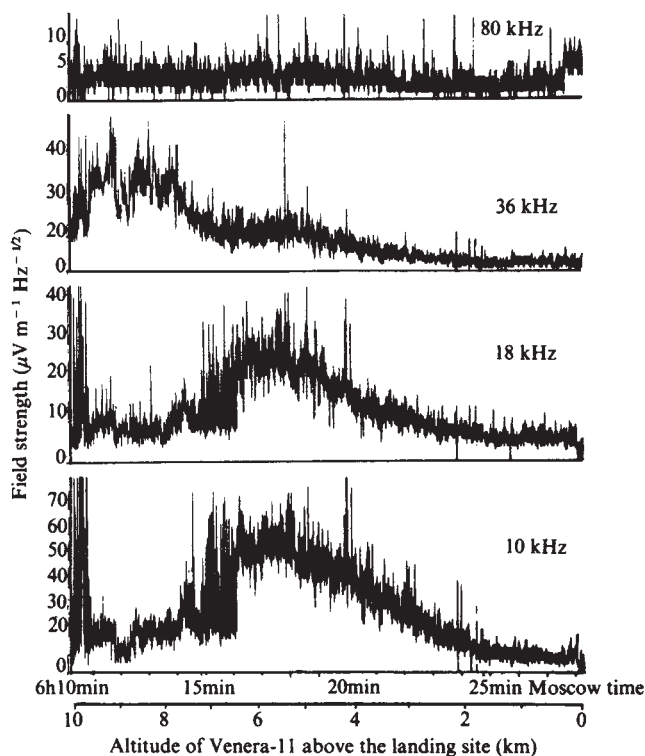


Fig. 4 Bursts recorded by the Venera 11 before the landing. It covers the altitudes 8–2 km. Immediately before the landing and on the surface the level of atmospheric radionoise was very low. The burst shows most prominence at 36 kHz.

Is something wrong with the binary pulsar?

W. Kundt

Institut für Astrophysik der Universität Auf dem Hügel 71,
D-5300 Bonn, FRG

The binary pulsar PSR1913 + 16 is at present the most interesting object for measuring general relativistic effects¹ such as the periastron advance, emission of gravitational waves, and gravitational redshift. It also offers unique possibilities for studying the properties of one or two neutron stars. Yet several of the observed effects could be simulated by an extended secondary, and recent optical measurements^{2,3} have suggested that such is the case, at the 97% confidence level. Here I shall argue that the optical source seen at the position of PSR1913 + 16 may be the pulsar sphere, that is, a region of confined extremely relativistic plasma glowing in the near IR, and that the secondary is another neutron star (as preferred by most authors). Therefore, the optical coincidence need not cause us to doubt the published general relativistic analysis of this system. I shall also comment on the binary pulsar's likely history, and on the possible presence of a second binary pulsar in Cas A.

At its present spindown rate¹ of $\dot{P} = 0.864 \times 10^{-17}$ and period $P = 0.0590 \text{ s}$, the binary pulsar loses rotational energy at a rate

$$-\dot{E} = -I\Omega\dot{\Omega} = 10^{33} I_{44.8} \text{ erg s}^{-1} \quad (1)$$

where $\Omega := 2\pi/P$. I is the moment of inertia, and

$I_{44.8} := I/10^{44.8} \text{ g cm}^2$ equals unity for the Crab pulsar⁴. Where does this power go?

In the case of the Crab, the pulsar is thought⁵ to illuminate the surrounding supernova remnant; that is, a substantial fraction of its rotational energy losses is thought to be converted into magnetobremstrahlung of the Crab nebula. More quantitatively⁴, half of the corresponding power may be pumped into the 30-Hz wave bath, and post-accelerate the expanding filaments as well as heat the lowest-energy relativistic electrons. The other half of that power is thought to be injected into the nebula in the form of highly relativistic electrons and positrons, with Lorentz factors γ in the range $3 \times 10^5 \leq \gamma \leq 10^7$, distributed according to $N_\gamma d\gamma \sim \gamma^{-2.2} d\gamma$. In the nebular electromagnetic field of strength $\approx 10^{-3} \text{ G}$, these particles lose their energy via magnetobremstrahlung whose power peaks at the frequency

$$\nu_B = eB_\perp \gamma^2 / \pi m_e c = 6 \times 10^{15} \text{ Hz } B_{-3} \gamma_6^2 \quad (2)$$

that is, in the UV.

If the Vela pulsar has the same moment of inertia as the Crab pulsar, some 10% of its spindown power escape as pulsed γ rays^{6,7}, between 0.3 and 3 GeV. Luminosities at radio frequencies⁸ and X-ray energies⁹ are about 10 times smaller than at γ -ray energies. In analogy to the Crab, more than half of the spindown power is expected to be converted into magnetobremstrahlung (= synchrotron, or synchro-Compton radiation), probably at IR frequencies. Because of the large angular extent of the source, such a flux is difficult to measure. The magnetic fields in Vela X may be of the order⁹ of $10^{-4.5 \pm 0.5} \text{ G}$, so that the observed parts of the spectrum and equation (2) imply injected Lorentz factors in the interval $10^5 \leq \gamma \leq 10^8$.

The binary pulsar has a period between those of the Crab and Vela pulsars, and an inferred surface field strength

$$B_s = (3c^3 |\dot{E}| / 2R^6 \Omega^4)^{1/2} = 1.8 \times 10^{10} \text{ G} \quad (3)$$

which is more than 10^2 times weaker. Here a standard neutron star radius $R = 10 \text{ km}$ has been assumed, and a perpendicular magnetic dipole field responsible for the radio pulses. The binary pulsar's spindown age, $t = P/2\dot{P} = 1.1 \times 10^8 \text{ yr}$ may be an overestimate of its true age; but t is certainly a measure of the pulsar's present ageing rate or lifetime. If it were a large overestimate of its true age, by a factor N , we should expect, on statistical grounds, to see N more binary pulsars in the sky.

As supernova remnants containing pulsars (not shrouded neutron stars) seem to disperse within little more than 10^4 yr , the binary pulsar will presently find itself surrounded by typical interstellar matter, with a static pressure $p_{\text{stat}} \leq 10^{-12} \text{ dyn cm}^{-2}$, and a hydrogen density $n = T_4^{-1} \text{ cm}^{-3}$. Its relativistic wind will be confined by this matter, and run into an inner shock at a distance r_i where the wind pressure $|\dot{E}|/4\pi r_i^2 c$ meets the cavity pressure p_{cav} . In a steady-state flow, p_{cav} equals the ram pressure

$$\rho v^2/2 \leq 10^{-9} n_0 v_{7.5}^2 \text{ dyn cm}^{-2} \quad (4)$$

of the ambient medium, where $v = 10^{7.5} \text{ cm s}^{-1}$ is the assumed relative velocity brought about by two successive supernova explosions, see equation (8), and the inner cavity radius is

$$r_i = (|\dot{E}|/4\pi c p_{\text{cav}})^{1/2} \geq 2 \times 10^{15} p_{-9}^{-1/2} \text{ cm} \quad (5)$$

The cavity must have a larger transverse than longitudinal extent for a large pressure anisotropy $p_{\text{ram}}/p_{\text{stat}} \gg 1$ of the confining surroundings. Its size is controlled by the condition that the pulsar's particle supply must leave through its outer edge at a speed v_{esc} typical for the (much heavier) ambient medium, which implies $r_0/r_i \approx (c/v_{\text{esc}})^{1/2} \leq 30$ for the outer radial scale r_0 , or

$$r_0 \leq 10^{17} p_{-9}^{-1/2} \text{ cm} \quad (6)$$

The magnetic field inside the cavity will at least reach pressure balance strength $B = 2 \times 10^{-4} p_{-9}^{-1/2} \text{ G}$. This value is even likely to be a large underestimate because the cavity boundary is strongly Rayleigh–Taylor unstable, and intruding filaments of interstellar matter will be wrapped into much stronger magnetic fields¹⁰, reminiscent of the magnetic knots in Cas A. The extremely relativistic electrons traversing the cavity have a synchrotron lifetime

$$t_{\text{syn}} \approx (m_e^5 c^9 / e^7 B^3 \nu)^{1/2} = 0.8 \times 10^3 \text{ yr } B_{-3.5}^{-3/2} \nu_{14}^{-1/2} \quad (7)$$

if radiating at the frequency ν in a field of $10^{-3.5} \text{ G}$. At a typical gyrocentre speed of $10^{7.5} \text{ cm s}^{-1}$, they therefore lose a significant fraction of their energy over a distance of several 10^{17} cm $B_{-3.5}^{-3/2} \nu_{14}^{-1/2}$. For an effective cavity radius of 10^{17} cm , particles may traverse some $3 \times 10^{17} \text{ cm}$ before they leave the vicinity of the pulsar, and are dragged along by the wake into regions of reduced pressure and much lower synchrotron brightness. Consequently, from a distance of $d = 5 \text{ kpc}$ the pulsar sphere is expected to look like a source of angular diameter $\leq 2.8 \text{ arc s}$ $d_{22.2}^{-1} p_{-9}^{-1/2}$, whilst the reported 'star' has an unresolved size of $\leq 2 \text{ arc s}$ in the deep red band³. (See also ref. 11 which excludes pulsations at the 1% level.) I propose that the reported star is in fact the glowing pulsar sphere.

This interpretation is strengthened by the fact that from the observations by Kristian *et al.*² and Crane *et al.*³ one can derive an $R - V$ spectral index $\alpha = -2.6$, ($S_\nu \sim \nu^\alpha$), very unlike that of a hot star; of course, α carries a large uncertainty. The detected flux of $1.6 \times 10^{-5} \text{ Jy}$ at $0.68 \times 10^{-4} \text{ cm}$ corresponds to an R -luminosity of $L_R = 2 \times 10^{32} \text{ erg s}^{-1} = 0.2 |E_{\text{PSR}}|$, in agreement with the model. Note that an additional power input of $\pi r_0^2 \rho v^3/2 = 5 \times 10^{30} \text{ erg s}^{-1}$ $r_{17}^2 v_{7.5}^3$ is expected from the ram pressure exerted against the interstellar medium, with voltages of the order of $2 r_0 e B_\infty v/c = 10^{12} \text{ eV } B_{-5} v_{7.5}$. This power is, however, small unless v exceeded 10^3 km s^{-1} .

The proposed identification of the binary pulsar's sphere with the starlike optical source near $(l, b) = (49.97^\circ, 2.1^\circ)$ is based on an assumed distance $\geq 5 \text{ kpc}$ corresponding to the dispersion measure $171.64 \text{ pc cm}^{-3}$. If the pulsar's rays crossed (E. Schröfer, personal communication) the Strömgren sphere of the star at $(l, b) = (49.6^\circ, 0.25^\circ)$ which has been tentatively classified as OV, the dispersion measure could be a large overestimate of distance, and the binary pulsar could be as near as 0.8 kpc . In view of the small observed variability (L. A. Fowler, personal communication) of the dispersion measure [$< 0.6 \times 10^{-4}$], however, this possibility is not very likely.

Are the binary pulsar's properties exceptional? Its low inferred surface field strength (equation (3)) has sometimes led to the conjecture¹² that it is the older of the two hypothetical neutron stars, and that it was born with a much higher magnetic field. There is, however, as yet no conclusive evidence that neutron star magnetic fields do decay on a time scale of 10^8 yr ; see refs 13–15. If pulsar dipole fields did decay, they would have to do so with a rather wide range of decay times, between $\sim 10^7$ and 10^{10} yr . (But magnetic field decay may have taken place in the—probably $\sim 10^{10} \text{ yr}$ old—X-ray bursters¹⁶.) But there is significant evidence that binary neutron stars are spun down by the wind of their companion star^{17,18} to periods which are in most cases too long to make them visible as pulsars. The older of the two neutron stars may hence be spinning too slowly to be detected. Why then does the younger neutron star lack a strong magnetic field?

The possible evolutionary histories of the binary pulsar have been carefully discussed elsewhere^{12,19,20}. In view of the more recent data, the most convincing configuration before the last supernova explosion in the system is a neutron star in close circular orbit around a He star. One can then calculate the presupernova parameters of the system under the idealising assumptions of: (1) an initially circular relative orbit; (2) an instantaneous and symmetrical explosion; and (3) vanishing momentum transfer to the companion during the explosion. The condition of coincident initial conditions at periastron yields¹²

$$\begin{aligned} (M_1^- - M_1)/M &= \varepsilon &= 0.617 \\ M_1^- &= M_1 + \varepsilon(M_1 + M_2) = 3.0 M_\odot \\ a^- &= (1 - \varepsilon)a &= 0.747 \times 10^{11} \text{ cm} \\ \omega^- &= [(1 + \varepsilon)/(1 - \varepsilon)^3]^{1/2} \omega = 2\pi/1.44 \text{ h} \\ v_{\text{rec}} &= \varepsilon(1 - \varepsilon^2)^{-1/2} (M_2/M)(GM/a)^{1/2} \\ &= 1.8 \times 10^7 \text{ cm s}^{-1} \end{aligned} \quad (8)$$

where superscript ‘-’ denotes presupernova values, ε = numerical eccentricity, a = semi-major axis, ω = orbital angular velocity, v_{rec} = recoil velocity, and $M_1 = (1.3 \pm 0.15) M_\odot$, $M_2 = (1.5 \pm 0.15) M_\odot$ are the best estimates at present (L. A. Fowler, personal communication) for the masses, with $M := M_1 + M_2 = 2.83 M_\odot$. Obviously, the system could not get disrupted in the explosion because there was not enough mass in the shell. The ‘missing mass’ in the shell may have been evaporated by the closely orbiting (first) neutron star¹². Hence an exceptional property of the binary pulsar (when compared with other pulsars) is the small separation of its progenitor system, which in turn may have resulted from an exceptionally small mass ratio of the massive binary from which it descended²¹.

I shall now argue that a close presupernova orbit is likely to imply a small magnetic dipole moment of the resulting neutron star, and may hence be the only exceptional property of the binary pulsar. The argument follows Ruderman and Sutherland²² in that the neutron star inherits its magnetic moment from the convective core of its progenitor star, whose equipartition magnetic flux decreases with decreasing convective motion and hence with decreasing spin. But a presupernova binary separation of one solar radius gives rise to such a strong tidal couple that near-synchronous rotation is enforced within²³

$$\tau_1^{\text{synch}} \approx \frac{\alpha}{K} \left(\frac{M_1}{M_2} \right)^3 \left(\frac{a^-}{R_1} \right)^9 \frac{1}{\omega^-} \approx 10^{-2} \text{ yr} \left(\frac{a^-}{R_1} \right)^9, \quad (9)$$

where $\alpha/K \approx 8$, and R_1 is the radius of the helium star (through atmosphere of which the neutron star orbits). A spin period of the core of more than 1 h is still short enough to give rise to a fast pulsar, but probably not fast enough to drive a strong dynamo²².

Why may the older of the two neutron stars have a higher mass? It is tempting to speculate that it has accreted some $0.1 M_\odot$ throughout several Myr from the atmosphere of its companion.

The binary pulsar, therefore, (still) allows us to think of neutron stars as objects with standard properties produced by ideal supernovae. All that may be special about it is the small separation of its progenitors, which prevents disruption, implies a small magnetic moment, a low luminosity, and high space velocity (imparted both by the first and second supernova explosion).

An even larger space velocity must have been imparted to the (light) supernova shell blown off when the binary pulsar was born (due to momentum conservation). As I have argued previously¹⁰, a light supernova shell [$< 3 M_\odot$] with a high space velocity [of $(165 \pm 15) \text{ km s}^{-1}$] is Cas A. Cas A is exceptional for its radio brightness: Tammann²⁴ estimates that at most 2% of all supernova remnants are similar to it. Its recoil partner should presently be located at (23 h 21 min $(12 \pm 0.5) \text{ s}$, $58^\circ 32' (17 \pm 1)''$). As no (strong) pulsar nor bright star has been found in this region, a (weak) binary pulsar is a likely candidate. It could, perhaps, be the third²⁵ binary system in which emission of gravitational waves is measurable.

I thank B. Eckart for discussions, and N. Holloway for the manuscript.

Received 30 October 1979; accepted 4 February 1980.

1. Taylor, J. H., Fowler, L. A. & McCulloch, P. M. *Nature* **277**, 437–440 (1979).
2. Kristian, J. & Westphal, J. A. *IAU Circ. No.* 3242 (1978).
3. Crane, P., Nelson, J. E. & Tyson, J. A. *Nature* **280**, 367–370 (1979).
4. Kundt, W. & Krotscheck, E. *Astr. Astrophys.* (in the press).
5. Rees, M. J. & Gunn, J. E. *Mon. Not. R. Astr. Soc.* **167**, 1–12 (1974).
6. Kanbach, G. et al. *Proc. 12th ESLAB Symp.* **21** (ESA SP-124, 1977).
7. Lichti, G. G. et al. *Advances in Space Exploration* Vol. 7 (Pergamon, Oxford, 1980).
8. Weiler, K. W. & Panagia, N. *Astr. Astrophys.* (submitted).
9. Smith, A. *Mon. Not. R. Astr. Soc.* **182**, 39P–46P (1978).
10. Kundt, W. *Ann. N.Y. Acad. Sci.* (in the press).
11. Nather, R. E., Robinson, E. L., Van Citters, G. W. & Hemenway, P. D. *Astrophys. J.* **211**, L125–L127 (1977).
12. Smarr, L. L. & Blandford, R. *Astrophys. J.* **207**, 574–588 (1976).
13. Jones, P. B. *Nature* **262**, 120–121 (1976).
14. Flowers, E. & Ruderman, H. A. *Astrophys. J.* **215**, 302–310 (1977).
15. Kundt, W. *Naturwissenschaften* **64**, 493–498 (1977).
16. Joss, P. C. *Astrophys. J.* **225**, L123–L127 (1978).
17. Van den Heuvel, E. P. J. *Ann. N.Y. Acad. Sci.* **302**, 14–35 (1977).
18. Holloway, N., Kundt, W. & Wang, Y.-M. *Astr. Astrophys.* **70**, L23–L26 (1978).
19. Flannery, B. P. & Van den Heuvel, E. P. J. *Astr. Astrophys.* **39**, 61–67 (1975).
20. Webbink, R. F. *Astr. Astrophys.* **41**, 1–8 (1975).

21. Vanbeveren, D. & Conti, P. S. *Astr. Astrophys.* (submitted).
22. Ruderman, M. A. & Sutherland, P. G. *Nature phys. Sci.* **246**, 93 (1973).
23. Press, W. H., Wiita, P. J. & Smarr, L. L. *Astrophys. J.* **202**, L135–L137 (1975).
24. Tammann, G. *Ann. N.Y. Acad. Sci.* **302**, 61–80 (1977).
25. Ritter, H. *Nature* **281**, 199–200 (1979).

Magnetic energy conversion, magnetospheric substorms and solar flares

S.-I. Akasofu

Geophysical Institute, University of Alaska, Fairbanks, Alaska 99701

An explosive conversion of magnetic energy has long been suggested as the cause of various electrodynamic processes in astrophysical conditions, such as auroral magnetospheric substorms, solar flares, stellar flares and some galactic scale processes^{1–6}. It is generally believed that the magnetic energy is stored in a certain region before the onset of those phenomena and that the conversion takes place spontaneously or by triggering. Here it is suggested that the magnetospheric substorm is a direct consequence of increasing power generated by the solar wind–magnetosphere dynamo, rather than of a hypothetical sudden conversion of the stored magnetic energy.

The magnetospheric substorm has been thought to be the manifestation of a sudden conversion of the magnetic energy stored in the magnetotail before substorm onset. It has been believed that solar flares arise from an explosive conversion of the magnetic energy stored in the coronal level above a large sunspot group. Among these and stellar situations, the magnetosphere is the only region which can be ‘diagnosed’ systematically by *in situ* measurements of plasma quantities, as well as of electric and magnetic fields, providing the grounds to test the hypothesis of a sudden magnetic energy conversion. The solar wind (a magnetised plasma flow) and the magnetosphere (a magnetised celestial body) constitute a dynamo. The dynamo-induced current flows in various parts of the magnetosphere, including the magnetotail and the ionosphere. Therefore, by comparing time variations of the input energy flow into the magnetosphere, the amount of the magnetic energy in the magnetotail and the energy consumed by the magnetosphere, one should be able to identify what processes are involved in magnetospheric substorms.

We have recently found a particular solar wind quantity ε which correlates well with the rate of the total energy $\bar{U}$ consumed by the magnetosphere⁷. It is given by:

$$\varepsilon = VB^2 \sin^4 \left(\frac{\theta}{2} \right) l_0^2 \quad (1)$$

where V (cm s^{-1}) is the solar wind speed; B (G) is the magnitude of the interplanetary magnetic field; θ is the polar angle of the IMF projected onto a plane perpendicular to the Sun–Earth line; l_0 is constant ($= 7R_E$); and $\bar{U}$ is the sum of the rate of the Joule heat production in the ionosphere $\bar{U}_i$, the rate of the ring current particle injection $\bar{U}_R$, and the rate of auroral particle injection $\bar{U}_A$.

Furthermore, it has been shown⁸ that ε is equal to the power P generated by the solar wind–magnetosphere dynamo, which is given by⁹

$$P = \frac{1}{4\pi} (B_{\text{tan}} B_{\text{norm}}) VS \quad (2)$$

where B_{tan} and B_{norm} denote the tangential and normal components of the magnetic field at the magnetopause and S denotes the surface area of the magnetotail.

It is thus of great interest to examine the relationship among the solar wind–magnetosphere energy coupling function (ε), the

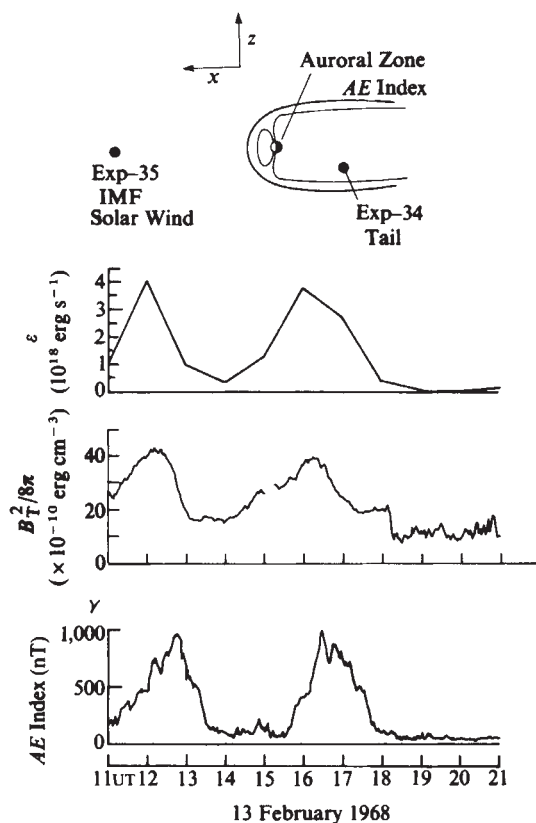


Fig. 1 Comparison of the solar wind-magnetosphere energy coupling function ϵ , the magnetic energy density $B_T^2/8\pi$ in the high latitude lobe in the magnetotail, and the magnetospheric substorm index AE . These data were obtained at Explorer 35 and 34 satellites and auroral zone stations, respectively, and their locations with respect to the magnetosphere are shown in the insert.

magnetic energy density ($B_T^2/8\pi$) in the magnetotail, where B_T denotes the magnetotail field in the high latitude lobe, and the substorm index $AE(\text{nT}) \approx \bar{U}_1/10^{15} \text{ erg s}^{-1}$ (ref. 10). Figure 1 shows the relationship between these three quantities monitored by the Explorer 34, 35 satellites and auroral zone magnetic observatories, respectively. This particular magnetotail event was chosen because it has been well documented¹¹. For the relative location of the satellites with respect to the magnetosphere, see the insert in Fig. 1.

Figure 1 shows that the three quantities, ϵ , $B_T^2/8\pi$ and AE vary roughly in harmony. Note particularly the occurrence of two substorms which are clearly indicated by an increase of the AE index; the corresponding auroral activity was well recorded in Alaska. If one assumes that the magnetospheric substorm is produced by an explosive conversion of the magnetic energy stored in the magnetotail before substorm onset, $B_T^2/8\pi$ should begin to decrease rapidly at $T=0$, namely at ~ 11.00 UT and ~ 15.30 UT, respectively. On the contrary, $B_T^2/8\pi$ increased with ϵ and the AE index at least during the main epoch (11.00–12.10 UT and 15.30–16.20 UT) of the expansive phase.

An important implication of this result is that the development of magnetospheric substorms is a direct consequence of increasing ϵ and the solar wind-magnetosphere dynamo power P . As the power is increased, both the magnetotail (solenoidal) current and the auroral electrojet are enhanced, as can be seen in the corresponding increase of $B_T^2/8\pi$ and AE . Note, however, that we assume here that the total volume of the magnetotail is not reduced appreciably. It has already been shown that ϵ and the AE index are related statistically by $AE(\text{nT}) \approx \epsilon/(10^{16} \text{ erg s}^{-1})$ for $\epsilon < 10^{19} \text{ erg s}^{-1}$ (ref. 10) although in this particular case, $AE(\text{nT}) \approx 2\epsilon/(10^{16} \text{ erg s}^{-1})$. Auroral activity manifests typical features of the substorm when the accompanying magnetic variations exceed about 100 nT (ref. 12); so that substorm activity is associated with $\epsilon \geq 10^{18} \text{ erg s}^{-1}$.

Note that although hourly values are used in computing ϵ , there seems to be an appreciable time delay between $\epsilon(t)$ and $AE(t)$. Such a delay can easily be expected for the magnetospheric circuit, as its inductance is known to be of the order of ~ 100 –500 H. Since the total ionospheric resistance is of the order of 0.1Ω , the time constant of the system is of the order of ~ 17 –85 min. Thus, the magnetosphere cannot respond instantly and fully to changes of ϵ . Any storing of energy should occur during the period when the dynamo power is increasing, not before substorm onset.

On the basis of Fig. 1 and similar events, there is little doubt that the magnetospheric substorm develops when the solar wind-magnetosphere dynamo is being driven harder than during a quiet period. The magnetospheric substorm is not simply a process by which the total magnetic energy is stored before its onset and then is explosively consumed. Further, it begins to subside as ϵ decreases, not because stored energy is consumed. Therefore, this clarifies why the magnetic energy conversion subsides at the end of substorms (as a large amount of magnetic energy is still available then or even during a quiet time).

The magnetospheric substorm may be said to be a direct consequence of increasing power generated by the solar wind-magnetosphere dynamo above $\epsilon > 10^{18} \text{ erg s}^{-1}$. Part of the increased current is channelled to the ionosphere along magnetic field lines. There are at least two important consequences of the increased field-aligned currents. The first is that the field-aligned currents develop a particular potential structure (called the V-shaped potential)¹³. As a result, the current-carrying electrons are accelerated downward and produce the aurora as they collide with polar upper atmosphere particles. As auroral electrons are accelerated to a few kilovolts, the auroral arc will suddenly brighten. An increased ionisation and conductivity might have a positive nonlinear effect¹⁴. The second consequence of the increased field-aligned currents is the generation of the auroral electrojet; a large increase of the Joule heat production (responding to a large increase of the power generation) can take place.

It may perhaps be premature to generalise this finding to other phenomena which have been suspected to be caused by a sudden magnetic energy conversion. On the other hand, the magnetosphere is the only region in which the hypothesis of an explosive magnetic energy conversion can be tested¹⁵. In this respect, solar flares have also been believed to be due to such an energy conversion process. Indeed, there is a close phenomenological similarity between magnetospheric substorms and solar flares; in fact, flare ribbons may be identified as solar auroras¹⁶. Although details are given elsewhere¹⁶, note that for a vertical wind of speed $V = 1 \text{ km s}^{-1}$ across horizontal magnetic field lines of $B = 500 \text{ G}$ in an area S of $10^5 \text{ km} \times 10^5 \text{ km}$, the dynamo power is given by $P = (B^2/8\pi) \times S \times V \approx 10^{29} \text{ erg s}^{-1}$. Therefore, as in the case of auroral substorm, in spite of its sudden onset, a solar flare could be a direct result of a photospheric dynamo process. In fact, it has been shown that a solar flare tends to occur in the vicinity of a newly emerging magnetic flux¹⁷, so that this emerging flux is likely to be an indication of the presence of such a dynamo process.

This work was supported by a grant from the NSF, Atmospheric Sciences Section, ATM77-26522.

Received 3 September 1979; accepted 17 January 1980.

- Švestka, Z. *Solar Flares* (Reidel, Dordrecht, 1976).
- Sweet, P. A. *Rev. Astr. Astrophys.* **7**, 149 (1969).
- Piddington, J. H. *Cosmic Electrodynamics* (Wiley, New York, 1970).
- Dungey, J. W. *Cosmic Electrodynamics* (Cambridge University Press, London, 1958).
- Parker, E. W. *Astrophys. J. Suppl.* **8**, 177, 1963.
- Akasofu, S.-I. *Physics of Magnetospheric Substorms* (Reidel, Dordrecht, 1977).
- Perreault, P. & Akasofu, S.-I. *Geophys. Res. Astr. Soc.* **54**, 547 (1978).
- Kan, J. R., Lee, L. C. & Akasofu, S.-I. *J. geophys. Res.* (submitted).
- Siscoe, G. L. & Cummings, W. D. *Planet. Space Sci.* **17**, 1795 (1969).
- Akasofu, S.-I. *Planet. Space Sci.* **27**, 425, 1979.
- Fairfield, D. H. & Ness, N. F. *J. geophys. Res.* **75**, 7032, 1970.
- Akasofu, S.-I., Kimball, D. S. & Meng, C.-I. *Ann. IGY*, Vol. XLV (Pergamon, Oxford, 1968).
- Kan, J. R., Lee, L. C. & Akasofu, S.-I. *J. geophys. Res.* **84**, 4305 (1979).
- Sato, T. & Holzer, T. E. *J. geophys. Res.* **78**, 7314 (1973).
- Fälthammar, C.-G., Akasofu, S.-I. & Allén, H. *Nature*, **275**, 285 (1978).
- Akasofu, S.-I. *Solar Phys.* (in the press).
- Rust, D. M. *Phil. Trans. R. Soc. A* **281**, 427 (1976).

The effect of compressible flow on anti-dynamo theorems

J. P. Todoeschuck & M. G. Rochester

Department of Physics, Memorial University of Newfoundland,
St John's, Newfoundland, Canada A1B 3X7

The 'anti-dynamo' theorems (ADTs) impose restrictions on the symmetry properties of the velocity and magnetic fields associated with sustained dynamo action in an electrically conducting body of fluid, such as the Earth's core. However, several of these ADTs depend on the assumption that the fluid flow is incompressible, and we show here that their validity is queried when compressible flow is considered. In particular, non-stationary axisymmetric dynamos may no longer be prohibited by Cowling's theorem. The high degree of axisymmetry of Saturn's magnetic dipole, revealed by the recent Pioneer 11 flyby, may be related to the planet's large compressibility.

The first and best-known of the ADTs is Cowling's¹ (extended by Backus and Chandrasekhar²), which proves the impossibility of supporting a stationary axisymmetric magnetic field $\mathbf{B}$ by axisymmetric fluid motions. The essential point in Cowling's argument is that motional induction cannot maintain $\mathbf{B}$ in the vicinity of the 'neutral' curve $\mathbf{B} = 0$, so the latter region is a sink of magnetic energy¹. If the requirement of stationarity is relaxed, this decay may be offset by diffusion into this region of field produced elsewhere in the conducting fluid body, and the proof of the ADT apparently fails. However, Cowling's theorem was later^{3,4} generalised, using a different proof based on global properties of the field, to forbid even non-stationary axisymmetric dynamos.

There seems to be a widespread impression that Lortz⁵ has shown that even the requirement of axisymmetry in the velocity field $\mathbf{v}$ can be abandoned without weakening Cowling's theorem, but Lortz's argument rests subtly on an assumption about axisymmetry of the electric field which in turn forces $\mathbf{v}$ to be axisymmetric.

Cowling's theorem has played an important part in the development of dynamo theory. It contributed to the scepticism with which several eminent physicists greeted Elsasser's pioneering studies in the 1940s and early 1950s. The widely felt suspicion that a more general and thoroughly prohibitive ADT might exist was not totally dispelled until 'existence proofs' were found independently by Herzenberg and by Backus in 1958. The necessity of avoiding the consequences of Cowling's theorem also decisively guided the Bullard-Gellman-Lilley-Gubbins and Braginskii approaches to a plausible terrestrial dynamo model.

Elsasser⁶ argued convincingly, and rigorous proofs were provided by Bullard and Gellman⁷ and by Backus⁸, that a toroidal $\mathbf{v}$ could not maintain dynamo action. This ADT has generally been interpreted as requiring $\mathbf{v}$ to have a non-zero radial component in a working dynamo. Analogous to this constraint is the impossibility of dynamo action when $\mathbf{v}$ lacks a component in one cartesian direction⁹. Cowling³ recognised that the toroidal $\mathbf{v}$, two-dimensional, and non-stationary axisymmetric ADTs are but special cases of a single class of theorems whose proofs use global properties of $\mathbf{B}$ and $\mathbf{v}$. Other ADTs have been conjectured¹⁰, but apparently the only other for which a rigorous proof exists is that prohibiting a stationary dynamo in a sphere when $\mathbf{v}$ is purely radial¹¹.

An almost universal assumption in dynamo theory is that the fluid flow can be treated as incompressible—that $\mathbf{v}$ is solenoidal, $\nabla \cdot \mathbf{v} = 0$. An exception to this is the paper by Namikawa and Matsushita¹¹, who remark that "a curl-free velocity field may be important for the dynamo maintenance of the magnetic field in the Earth's core". Also Krause and Roberts¹² find that, when

the electrical conductivity is low enough, finite compressibility can assist dynamo action in mirror-symmetric turbulence, but not enough to actually bring about regeneration.

As Smylie and Rochester¹³ have shown, the assumption of solenoidal flow is not permissible for large-scale radial motions in the stratified liquid core of the Earth. The relative importance of compression by transport through the pre-existing hydrostatic pressure field is measured by the dimensionless number $C = gL/c^2$, where g and c are representative values for gravity and compressional wave speed in the liquid core, and L is the radial length scale of the motion. Only for $C \ll 1$ can the flow be treated as solenoidal. For the Earth's core, $C \sim 0.1$ when $L \sim 10^6$ m. The neglect of compressibility may therefore be of questionable validity for dynamo mechanisms which do not depend on small-scale turbulence in the core.

Among the most interesting consequences of taking compressibility into account are those for the ADTs which rest on 'global' proofs. Cowling's original 'neutral-curve' argument¹ is not dependent on any assumption of solenoidal flow, so his theorem forbidding a strictly stationary axisymmetric dynamo is unaffected. The same is true of the radial- $\mathbf{v}$ ADT discovered by Namikawa and Matsushita¹¹. But the proofs^{3,4} of the three other ADTs, using global properties of the fields, fail when a lamellar part $\nabla \mathcal{L}$ is added to $\mathbf{v}$. Separate clear standard proofs for each of these ADTs have been collected recently by Moffatt¹⁴.

The 'uncurled' hydromagnetic induction equation for axisymmetric $\mathbf{B}$ and $\mathbf{v}$ leads to the equation

$$\frac{d}{dt} \int_{V_\infty} \frac{\chi^2}{2\lambda} dV + \int_{V_\infty} (\nabla \chi)^2 dV = \int_{V_\infty} \frac{\chi^2}{2\lambda} \nabla^2 \mathcal{L} dV \quad (1)$$

where λ is the magnetic diffusivity, V_∞ is all space, and $\chi = -r \sin \theta (\partial P / \partial \theta)$ is the flux function associated with the poloidal part of $\mathbf{B}$, $= \nabla \times (\nabla \times \mathbf{r} P)$. Here r and θ are the usual spherical polar coordinates. If the RHS of equation (1) vanishes, we have Moffatt's equation (6.24), from which it is clear that the poloidal magnetic field must ultimately decay. But this conclusion is no longer assured for flows such that $\nabla^2 \mathcal{L}$ is non-negative in some finite region of space. A similar result is obtained for the toroidal part of $\mathbf{B}$ by the appropriate modification of Moffatt's equation (6.27). If U is a representative value of the radial component of the lamellar part of $\mathbf{v}$, then $\nabla \cdot \mathbf{v} \sim UC/L$, and equation (1) indicates that the proof of this ADT is lost if $UC/L > 2\lambda/L^2$, that is if $CR_m > 2$, where $R_m = UL/\lambda$ is the magnetic Reynolds number. This requirement does not seem too difficult to meet in the Earth's core, and should be even easier to satisfy in an even more compressible planet such as Saturn. The possibility that non-stationary axisymmetric dynamos can be sustained by lamellar flow deserves investigation.

By definition a toroidal $\mathbf{v}$ is solenoidal and has no radial component. If the first of these conditions is relaxed as above, but with $\partial \mathcal{L} / \partial r = 0$, Moffatt's equation (6.41) is replaced by the equation

$$\frac{d}{dt} \int_V Q^2 dV + 2\lambda \int_{V_\infty} (\nabla Q)^2 dV = - \int_V Q^2 \nabla^2 \mathcal{L} dV \quad (2)$$

where $Q = \mathbf{r} \cdot \mathbf{B}$, and V is the volume of the conducting fluid (in which λ is supposed uniform). When its RHS vanishes, equation (2), together with Moffatt's equation (6.46) (which is similarly affected by lamellar flow), yields the ADT for toroidal $\mathbf{v}$. But clearly the usual interpretation (that a radial component of $\mathbf{v}$ is required for dynamo action) may be voided if sufficiently compressible flow is allowed. However, because of the large non-radial compression required to violate it, the proof of this ADT for non-radial $\mathbf{v}$ remains valid in the Earth's core, at least. The prohibition against dynamos driven by plane two-dimensional motion⁹ may be similarly withdrawn in principle by permitting compressible flow.

We conclude that the usual proofs of several ADTs are nullified by abandoning the requirement of strictly solenoidal flow. We do not yet know if this means that the theorems themselves are invalid for compressible flow—that alternative

proofs independent of the incompressibility assumption cannot (or that counter examples can) be found. We do suggest that allowing for the physically real compressibility of electrically conducting fluids in planets may considerably relax (or remove) the constraint imposed by one of the ADTs, and permit the construction of non-stationary axisymmetric dynamos. For evidence that this result is not of purely academic interest we cite: (1) the likelihood, based on the known history of the Earth's magnetic field, that physical dynamos are essentially nonstationary on a wide range of time scales; (2) the recent discovery by Pioneer 11, that Saturn's magnetic dipole is highly axisymmetric¹⁵.

Received 11 January; accepted 28 January 1980.

1. Cowling, T. G. *Mon. Not. R. astr. Soc.* **94**, 39–48 (1933).
2. Backus, G. & Chandrasekhar, S. *Proc. natn Acad. Sci. U.S.A.* **42**, 105–109 (1956).
3. Cowling, T. G. *Q. J. Mech. appl. Math.* **10**, 129–136 (1957).
4. Braginskii, S. I. *Sov. Phys. JETP* **20**, 726–735 (1964).
5. Lortz, D. *Phys. Fluids* **11**, 913–915 (1968).
6. Elasser, W. M. *Phys. Rev.* **69**, 106–116 (1946).
7. Bullard, E. C. & Gellman, H. *Phil. Trans. R. Soc. A* **A247**, 213–278 (1954).
8. Backus, G. *Ann. Phys.* **4**, 372–447 (1958).
9. Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids*, 121–122 (Cambridge University Press, 1978).
10. Childress, S. *Théorie Magnétohydrodynamique de l'Effet Dynamo* (Institut Henri Poincaré, Paris, 1968).
11. Namikawa, T. & Matsushita, S. *Geophys. J. R. astr. Soc.* **19**, 395–415 (1970).
12. Krause, F. & Roberts, P. H. *Mathematika* **20**, 24–33 (1973).
13. Smylie, D. E. & Rochester, M. G. *Eos* **60**, 753 (1979).
14. Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids*, 113–122 (Cambridge University Press, 1978).
15. Smith, E. J. *et al. Eos* **60**, 865 (1979).

Anomalously high uranium contents in the sediment under Galapagos hydrothermal mounds

Claude Lalou & Evelyne Brichet

Centre des Faibles Radioactivités, CNRS-CEA, 91190 Gif sur Yvette, France

The geochemical behaviour of uranium and thorium in metalliferous sediments and hydrothermal deposits has been widely studied and the main results have been summarised by Boström and Rydell¹. These isotopes may be used to clarify how the metal-rich solutions are introduced into sediment cover and seawater. Using radiochemistry followed by α spectrometry, we have measured uranium concentrations as high as several hundred p.p.m., which must clearly be associated with ocean ridge thermal activity, in sediments interbedded between the basaltic basement and the green hydrothermal mud at DSDP Site 424. These high uranium concentrations indicate the path followed by the hydrothermal fluid which, debouching at the sediment–water interface, formed the green mud.

Very high uranium concentrations have been reported in East Pacific Rise (EPR) metalliferous sediments by Boström and Rydell¹ who found, on a carbonate-free basis, average value of 38 p.p.m. (maximum 62 p.p.m.). In the metalliferous sediments of Red Sea Brines, a maximum value of 31.3 p.p.m. has been found². The highest value presently reported for uranium concentration is 150 p.p.m. (CaCO₃-free basis) for a level in core GS-7202-35 (14°47' S, 113°30' W)³. However, in metalliferous sediments near the basaltic basement at DSDP sites 37, 38 and 39, only a weak enrichment in uranium was found⁴. The high concentrations are generally attributed to the scavenging of seawater uranium by the precipitation of iron hydroxides when hydrothermal solutions debouch in the seawater as the $^{234}\text{U}/^{238}\text{U}$ ratio in these deposits is close to the isotopic ratio of seawater. When higher isotopic ratios are found, they are considered to be due to hydrothermal leaching from basalt without subsequent important mixing with seawater uranium; the uranium content is generally relatively low⁵.

During DSDP Leg 54, four holes have been drilled at Site 424, an area of mounds previously described^{6–9} and of presumably hydrothermal origin. A relatively detailed description of the drill-holes (424, 424 A, B and C) drilled about 22 km south of the spreading axis (00°35' N, 86°07' W) is given by Hékinian *et al.*¹⁰ and Natland *et al.*¹¹. Some lithological data are available^{10,11}: the mean sedimentation rate in the area, calculated from Site 425 is 4.5 cm kyr⁻¹, if a constant and regular sedimentation rate is accepted for the 83-m thick sediment blanket covering a 1.8 Myr-old basement.

The schematic lithological sequence in holes 424, 424 A and B is, starting at the basaltic basement:

(1) a layer of pelagic sediment (referred to here as the basal sediment), essentially Foraminifera and nannocalcareous ooze, with thicknesses of 18.5, 13 and 13.5 m at each hole respectively,

(2) a relatively thick (15 m) layer of hydrothermal deposit formed with a green clay-rich material. This clay is essentially formed with nontronite,

(3) capping the whole: in holes 424 and 424 A drilled directly on the mounds a Mn oxide-rich layer is found, while in hole 424 B, between the mounds, the iron rich hydrothermal deposit is capped with calcareous sediments.

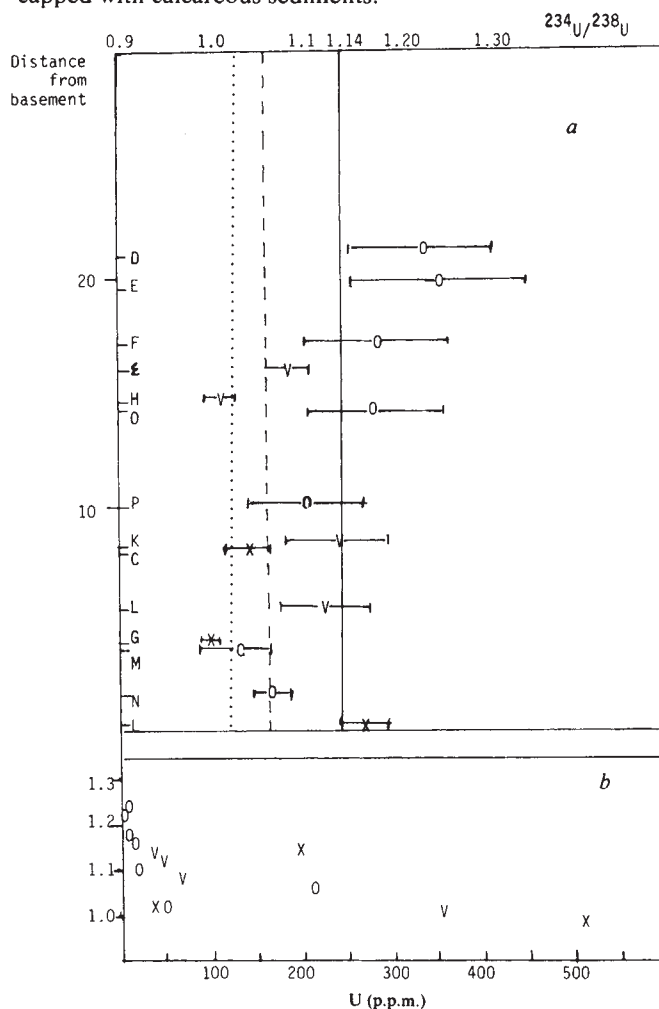


Fig. 1 a , $^{234}\text{U}/^{238}\text{U}$ activity ratio variations in the three drilled holes. Depths in the holes (m) have been arbitrarily calculated from the basaltic basement and are approximate values. Letters represent the laboratory number of the samples (see Table 1). O, values found in hole 424; V, values found in hole 424 A; X, values found in hole 424 B. Bars represent uncertainties calculated as one standard deviation (1σ) derived from counting statistics. Solid line indicates present seawater ratio (1.14); dashed line indicates the ratio for a seawater uranium deposited 300,000 yr ago (1.060); dotted line indicates the ratio for seawater uranium deposited 600,000 yr ago (1.026). b , $^{234}\text{U}/^{238}\text{U}$ activity ratio versus uranium concentrations (symbols as in a).

Table 1 Uranium and thorium concentrations, activity ratios and percentage CaCO_3

Hole	Lab. no.	DSDP no.	^{238}U (p.p.m.)	^{232}Th (p.p.m.)	$^{234}\text{U}/^{238}\text{U}^*$	$^{234}\text{U}/^{238}\text{U}^\dagger$	CaCO_3 (%)
424	D	2.5. 80. 81	3.62 ± 0.28	2.33 ± 0.32	1.233 ± 0.080	1.539	58
	E	2.6. 98. 99	5.77 ± 0.54	3.38 ± 0.42	1.250 ± 0.10	1.578	76
	F	3.2. 71. 72	6.13 ± 0.46	3.97 ± 0.48	1.180 ± 0.075	1.416	69
	O	3.3.113.114	10.78 ± 0.61	5.17 ± 0.68	1.177 ± 0.075	1.409	82
	P	3.6. 85. 86	14.94 ± 0.75	4.23 ± 0.65	1.106 ± 0.062	1.245	83
	M	4.4. 90. 91	46 ± 2	7.5 ± 1.2	1.028 ± 0.042	1.065	89
	N	4.5.130.131	207 ± 7	9.0 ± 1.5	1.066 ± 0.021	1.152	89
	ϵ	2.2.125.127	61.3 ± 1.6	6.5 ± 0.6	1.089 ± 0.021	1.208	88
424 A	H	2.3.109.110	344 ± 17	11.0 ± 1.4	1.010 ± 0.016	1.023	95
	K	3.1.139.140	33.0 ± 1.7	8 ± 1	1.135 ± 0.057	1.312	92
	L	3.3.101.102	43 ± 2	10.9 ± 1.7	1.121 ± 0.051	1.28	94
424 B	C	3.5. 20. 21	35.6 ± 1.9	3.7 ± 1.1	1.037 ± 0.024	1.086	79
	G	4.1. 95. 96	510 ± 27	5 ± 1	0.989 ± 0.07		88
	I	4.3. 80. 81	196 ± 9	8 ± 2	1.160 ± 0.029	1.37	93

* Activity ratio.

† Corrected for a 300,000 yr decrease of ^{234}U excess (see text).The uncertainties are one standard deviation (1 σ) derived from the counting statistics.

Natland *et al.*¹¹, using the age of the basement (about 0.6 Myr), the thickness of the basal sediment (about 15 m) and a sedimentation rate of about 5 cm kyr⁻¹, arrive at an age of 300,000 yr for the hydrothermal event. This conclusion is corroborated by the finding of two acoustic reflectors⁷ which can be attributed to this anomalous layer, the lowest one pinching on the crust, about 12 km from the Galapagos rift, on a crust about 300,000 yr old¹¹.

From the previous preliminary studies^{10,11} the two hydrothermal formations, green mud and manganese oxides, must have been deposited by two different events, the first one 300,000 yr ago, during which reducing conditions prevailed, the second one, recent and perhaps always active, in oxidising conditions.

We consider here the detailed study of the uranium and thorium content, and of the $^{234}\text{U}/^{238}\text{U}$ activity ratio in the basal sediment, deposited between 600,000 and 300,000 yr ago. The study of the distribution of these nuclides and of possible disequilibrium in the series in the hydrothermal mud, the manganese deposit and the upper sediment layer will be described elsewhere¹².

Uranium and thorium in the basal sediment were measured by α spectrometry, using either a solid state detector or a ionisation chamber, associated with a multichannel pulse height analyser. Uranium and thorium were separated and purified by co-precipitation, solvent extraction, ion exchange and final extraction with 1-(2'-thenoyl)-3,3,3-trifluoroacetone, followed by evaporation of the solution on stainless steel plates. As a spike, we used a ^{228}Th solution in secular equilibrium with ^{232}U . No measurements were made without the spike to verify that the equilibrium between natural ^{228}Th and ^{232}Th was achieved, because we only had about 1 g of sediment for each level and as ^{232}Th concentrations are so low that even if a disequilibrium were to exist, this would not greatly affect the results.

These sediments being very rich in Foraminifera and nannocalcareous ooze, which are low in uranium (0.1–0.13 p.p.m.)¹³, then dilute the authigenic or detrital uranium, the uranium and thorium concentrations are given in Table 1 on a dry and carbonate-free basis. Table 1 also shows the CaCO_3 content, the $^{234}\text{U}/^{238}\text{U}$ activity ratios, and the $^{234}\text{U}/^{238}\text{U}$ activity ratios corrected for a possible decrease during 300,000 yr.

A striking enrichment in uranium is seen, amounting to 500 p.p.m. in hole 424 B, core 4, section 1, 95–96 cm, but also affecting the basal sediment of the two other holes. In contrast, the green hydrothermal mud has a very low uranium content (0.2–0.9 p.p.m.), as does the recent sediment in hole 424 B (0.9–5 p.p.m., CaCO_3 free)¹².

In contrast ^{232}Th is depleted relative to normal pelagic clays (13 p.p.m. (ref. 14)), which may indicate that the non-carbonate

fraction contains very few detrital clays. Nevertheless, ^{232}Th concentrations are higher than in the green hydrothermal mud (about 1 p.p.m. (ref. 12)).

The first important point is that, in contrast with other metaliferous deposits, the uranium enrichment is not found in the hydrothermal deposit, but in the sediment below. Two hypotheses may explain this high uranium concentration in the sediment.

First, the uranium was concentrated from seawater in a reducing environment before the hydrothermal event which gave the green mud. In fact, in normal oxygenated marine environment uranium stays in solution, but high authigenic concentrations are encountered in shallow anaerobic sediments; concentrations amounting to 40 p.p.m. have been found in the organic-rich sediments of the Norwegian fjords¹⁵. Concentrations of the same order of magnitude associated with high molybdenum concentrations have been measured in a sediment core by Turekian and Bertine¹⁶ who interpret them 'as reflecting in part the formation of local ephemeral basins along the ridge, and in part the result of the high supply of organic material'. Bonatti *et al.*¹⁷ in core P 6702-59 (20°45' N, 85°20' W) have found an increase of uranium content with depth that they attributed to a change in redox potential and more recently, Mangini and Dominik¹⁸ have reported high values of uranium for eastern Mediterranean sapropels.

If this origin is retained for our samples, the seawater uranium was concentrated between 600,000 and 300,000 yr ago, period during which the basal sediment accumulated. If so, according to the 2.48 10^5 yr half life of ^{234}U , and with the assumption of closed system for uranium, the original $^{234}\text{U}/^{238}\text{U}$ activity ratio of seawater would have decreased from 1.14 (ref. 19) to 1.060 for uranium deposited 300,000 yr ago, and to 1.026 for the one deposited 600,000 yr ago.

Figure 1a shows the values of the measured ratios. Only three present a $^{234}\text{U}/^{238}\text{U}$ activity ratio compatible with the 600,000 yr value, and two with the 300,000 yr value. All the others are higher and five of them are even higher than the present seawater value.

However, this hypothesis would not explain why low $^{234}\text{U}/^{238}\text{U}$ ratios are associated with the highest uranium content (Fig. 1b). As Ku²⁰ has already shown, the postdepositional mobility of ^{234}U means that the hypothesis of closed system for uranium in sedimentary material is no more valuable. In fact, Mangini and Dominik¹⁸ have also found, in a 28-cm thick sapropel layer, the same inverse relation between $^{234}\text{U}/^{238}\text{U}$ ratio and uranium concentration, but, in this case, associated with a strong concentration gradient. They explain the variation of $^{234}\text{U}/^{238}\text{U}$ ratio by a half closed system, in which ^{234}U excess from seawater is in a closed system, while ^{234}U *in situ* produced

by radioactive decay is mobile. This mobile fraction is able to diffuse along the concentration slope, leading to an excess of ^{234}U in the low uranium material. As Kolodny and Kaplan²¹ have shown, in phosphorite the mobility of ^{234}U is due to the differential oxidation state of ^{234}U , $^{234}\text{U}/^{238}\text{U}$ being <1 in the U(IV) fraction and >1 in the U(VI) fraction.

In the case of Mediterranean sapropels, formed during anoxic periods, the uranium certainly originates in seawater. In hole 424 sediments, the presence of benthic Foraminifera (*Cassidulina laevigata*) indicates the presence of oxygenated waters at the sediment-water interface during deposition, so that the origin of uranium cannot be seawater.

Second, the concentration of uranium is of hydrothermal origin, its precipitation being due to the reducing conditions of the hydrothermal event. This is corroborated by the fact that the hydrothermal fluid was largely reducing as it has

given reduced iron deposits, even debouching in oxygenated seawater.

Hekinian *et al.*¹⁰ proposed two models for the mode of emplacement of these hydrothermal deposits. In the first model, there is a diffuse interaction with the sediment of an hydrothermal fluid from a deep-seated source. In the second model, the mounds are the narrow vents whose fluids do not interact with the sediment. From the uranium data, we conclude that the hydrothermal fluid has percolated through the sediment. Because of the steep gradient of uranium concentration, and the quantity of uranium mobilised in such a phenomenon, we suggest that the vents are not far from the mounds.

We thank Dr J. Labeyrie, J. C. Duplessy, R. Chesselet, L. Labeyrie and Dr K. K. Turekian for helpful comments. The NSF provided the samples. This study has been partly supported by DGRST grant 78-7-2197.

Received 7 November 1979; accepted 4 February 1980.

- Boström, K. & Rydell, H. *Coll. Intern. CNRS No. 289* (CNRS, Paris in the press).
- Ku, T. L. in *Hot Brines and Recent Heavy Metal Deposits in the Red Sea* (eds Ross, D. & Degens, E.) 512-524 (Springer, New York 1969).
- Rydell, H., Kraemer, T., Boström, K. & Joensuu, O. *Mar. Geol.* **17**, 151-164 (1974).
- Dymond, J. *et al. Bull. geol. Soc. Am.* **84**, 3355-3371 (1973).
- Rydell, H. S. & Bonatti, E. *Geochim. cosmochim. Acta* **37**, 2557-2565 (1973).
- Klitgorol, K. D. & Mudie, J. D. *Geophys. J. R. astr. Soc.* **38**, 563-586 (1974).
- Lonsdale, P. *Earth planet. Sci. Lett.* **36**, 92-110 (1977).
- Corliss, J. B., Lyle, M., Dymond, J. & Krane, K. *Earth planet. Sci. Lett.* **40**, 12-24 (1978).
- Corliss, J. B. *et al. Science* **203**, 4385, 1073-1083 (1979).
- Hekinian, R. *et al. Oceanol. Acta* **1-4**, 473-482 (1978).

- Natland, J. H. *et al. Science* **204**, 613-616 (1979).
- Lalou, C., Brichet, E. & Jehanno, C. (in preparation).
- Holmes, C. W., Osmond, J. K. & Goodell, H. G. *Earth planet. Sci. Lett.* **4**, 368-374 (1968).
- Ku, T. L. thesis, Columbia Univ. (1966).
- Veeh, H. H. *Earth planet. Sci. Lett.* **3**, 145-150 (1967).
- Turekian, K. K. & Bertine, K. K. *Nature* **229**, 250-251 (1971).
- Bonatti, E., Fisher, D. E., Joensuu, O. & Rydell, H. S. *Geochim. cosmochim. Acta* **35**, 189-201 (1971).
- Mangini, A. & Dominik, J. *Sediment. Geol.* **23**, 113-125 (1979).
- Ku, T. L. *A. Rev. Earth planet. Sci.* **4**, 347-379 (1976).
- Ku, T. L. *J. geophys. Res.* **70**, 3457-3474 (1965).
- Kolodny, Y. & Kaplan, I. R. *Geochim. cosmochim. Acta* **34**, 3-24 (1970).

Possible late Quaternary pingo remnants in central London

J. N. Hutchinson

Department of Civil Engineering, Imperial College, London SW7 2BU, UK

Active pingos are ice-cored mounds formed within permafrost¹. Actual and suspected remnants of such features have been reported widely¹⁻⁷ from former periglacial and glacial areas of the British Isles (Fig. 1)⁸⁻¹⁰. They are characteristically located in slope-foot situations with associated present-day springs or poor drainage⁴⁻⁶. Most of the observed features lie between the latitudes of The Wash and River Thames and are generally thought to have originated as open-system pingos¹¹ in the late Devensian. No pre-Devensian pingos have hitherto been recognised. Here I suggest that many of the anomalous, drift-filled, closed depressions in the surface of the rockhead (usually the London Clay) beneath some of the central London terraces, recently reviewed by Berry¹², are the partly eroded remains of former open-system pingos, dating both from the Devensian and from earlier cold stages of the Pleistocene.

These drift-filled depressions have been revealed by engineering works in central London from the late nineteenth century onwards¹². In their upper parts they seem to be of smooth contour and shaped like a wide funnel, oval in plan. In plan, the widths across their minor axes lie between about 20 and 150 m: the ratio of these to the dimensions of the major axes range generally from ~0.4 to 0.9. Their lower parts are less regular, narrower and steeper-sided (Fig. 2), in one instance featuring a buried vertical cliff of London Clay more than 12 m high. At least five of the reported depressions completely penetrate the London Clay. In these cases, the underlying stratum (generally the Woolwich and Reading Beds) has risen diapirically into the lower parts of the depression, typically by 6-12 m. In one of the several examples where the London Clay has apparently not been penetrated, a vertical tongue of this material at least 4 m high has diapiressed upwards into the base of the drift-filled depression. The maximum confirmed depth of the depressions below the general trend of the bench surface is ~25 m.

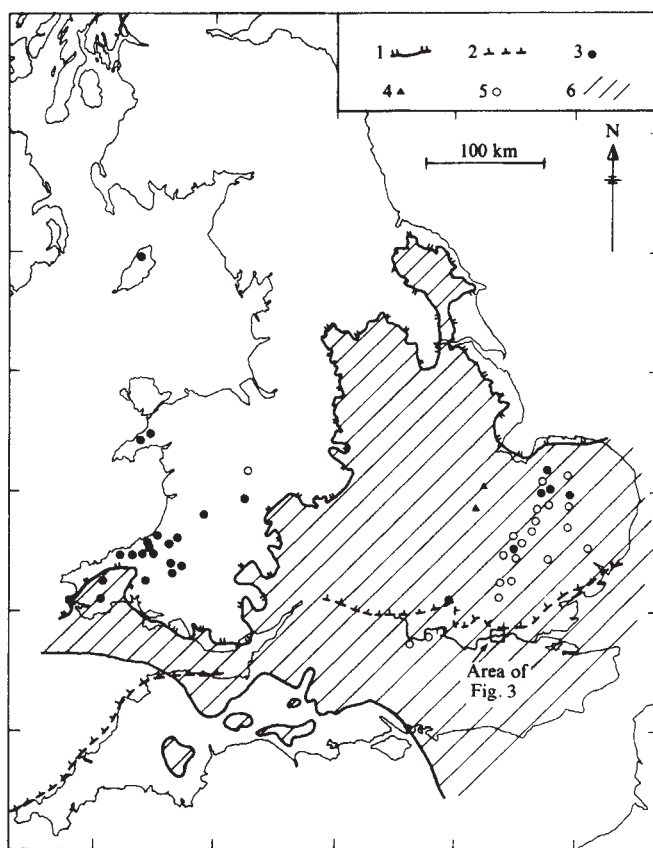


Fig. 1 Interim distribution of actual and possible pingo remnants in southern Britain, in relation to former maximum ice limits. (Acknowledgements to Committee for Aerial Photography, University of Cambridge, A. Taylor and J. Rose.) 1, Maximum southerly extent of Devensian ice⁸; 2, maximum southerly extent of pre-Devensian ice⁹; 3, localities of probable open-system pingo remnants in relief^{1-3,5,6}; 4, localities of possible closed-system pingos⁷; 5, localities where soil- or crop-marks indicate possible remnants of pingos or other ground ice features; 6, inferred distribution of extensive permafrost in the Devensian¹⁰.

It has been variously proposed that these features were formed by subsidence into solution cavities in the underlying Chalk¹³ (now generally discounted¹²), by some type of periglacial action^{12,14}, or by fluvial scour^{12,15}. The latter hypothesis has recently been reinforced by evidence for strong fluvial erosion during the late Devensian in stiff cohesive material in the channel of the River Gipping, Suffolk¹⁶. We accept that the anomalous depressions found in London have probably suffered major modification through fluvial scour, but do not consider that this mechanism can explain the origin of the features.

It is suggested here that many of these depressions originated as open-system pingos in the London Clay bedrock during the Devensian or earlier cold stages of the Pleistocene, deriving their water supply from artesian sub-permafrost or intra-permafrost groundwater through an unfrozen pipe or root. With the onset of interglacial or interstadial conditions, the collapsed pingo mounds would have been readily eroded away by the swollen rivers, to leave their roots exposed to local fluvial scour. This will have tended to enlarge and streamline at least the upper parts of the pingo roots, with corresponding replacement of some of the original infill by fluvial sediments. Some of the features may have been sheltered enough, or active sufficiently late, to suffer only minor modification.

The above hypothesis is supported by:

- (1) The general distribution of pingo remnants in Britain (Fig. 1), which indicates that conditions favouring the formation of open-system pingos may well have existed in the London area during the Devensian and will certainly have obtained there during earlier cold periods of the Pleistocene.
- (2) The inference, from the relation of the features to their infill, that the anomalous depressions in the London Clay were initiated in the "later stages of (late Quaternary) cold periods¹²", when the conditions of discontinuous permafrost usually associated with open-system pingos^{17,18} would have been likely to exist.
- (3) The observation that some of the depressions (such as, 3b)¹² have apparently been infilled partly by reworked London Clay^{12,19}, which may have been derived from the collapse and sloughing of a pingo mound.

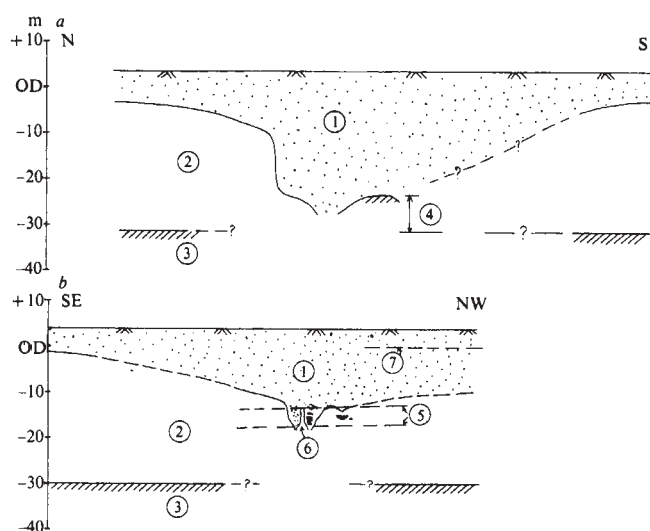


Fig. 2 Outline natural scale cross-sections of two of the anomalous depressions in central London, described by Berry¹²: *a*, N-S section of anomaly 1g, New Covent Garden, which penetrates the London Clay: 1, drift (chiefly sands and sandy gravels; sometimes clayey); 2, London Clay; 3, Woolwich and Reading Beds; 4, anomalous rise in level of junction of 2 and 3. *b*, NW-SE section of anomaly 2d, Vauxhall Grove, which appears not to penetrate the London Clay: 1-3, as for *a*; 5, Northbound tunnel of Victoria Line; 6, vertical tongue of London Clay; 7, general level of London Clay bench surface in the surrounding area.

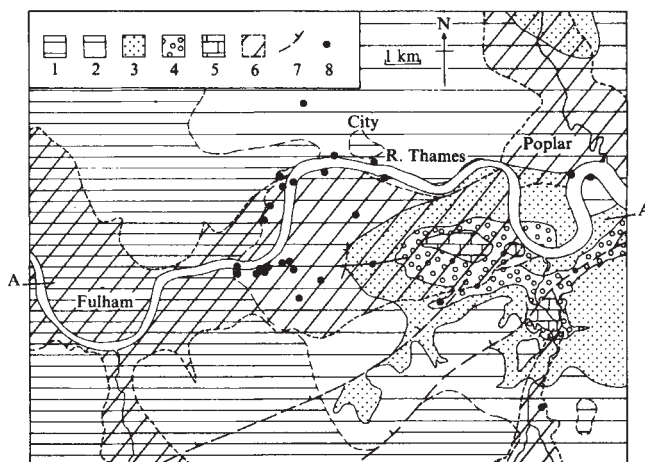


Fig. 3 Distribution of reported anomalous depressions in central London¹² in relation to the sub-crops of solid strata beneath the drift and to the estimated areas of former flowing artesian groundwater²⁰: 1, London Clay, greater than 35 m thick; 2, London Clay, less than 35 m thick; 3, Woolwich and Reading Beds; 4, Thanet Beds; 5, Upper Chalk; 6, former natural flowing artesian areas; 7, faults; 8, location of known anomalous depressions.

- (4) The distribution of the reported anomalous depressions in central London, which shows that: (i) Nearly all the depressions lie within, or close to, the former flowing artesian areas of this part of the London Basin²⁰ (Fig. 3). This suggests that groundwater conditions favouring open-system pingo formation^{11,21,22} are likely to have been present at the time when these features originated. (ii) The reported depressions occur mainly in the

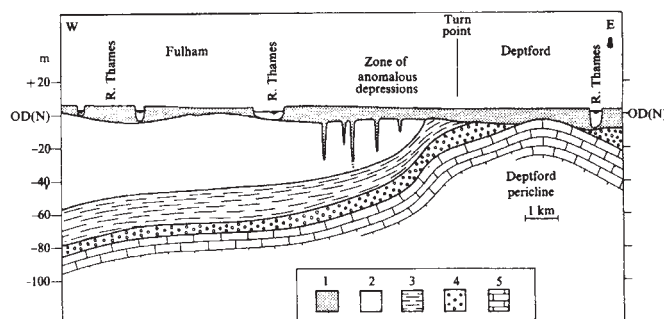


Fig. 4 Section A-A (Fig. 3) along the Thames Valley showing the relationship of the anomalous depressions to the feather-edge of the London Clay against the Deptford pericline: 1, Made ground and drift; 2, London Clay; 3, Woolwich and Reading Beds; 4, Thanet Beds; 5, Upper Chalk. Vertical exaggeration, $\times 50$.

feather edge of the London Clay stratum as it thins out against the inlier of Lower London Tertiaries and Chalk produced by the Deptford pericline (Fig. 4). In particular, they are restricted to the zone in which the present thickness of the London Clay is less than ~ 35 m (Figs 3 and 4). Bearing in mind that for open-system pingo formation a restricted supply of water is required^{11,21}, the absence of depressions from the artesian tracts where the London Clay is thicker may be the result of an inadequate upward groundwater flow. Conversely, their rarity in the artesian areas where the Lower London Tertiaries and the Chalk form the rock-head may indicate that in these localities the groundwater discharged too rapidly for pingo growth and

formed surface icings or naleds²² instead. (iii) The depressions show a propensity to cluster, generally in slope-foot or valley bottom situations; a characteristic that is entirely consistent with an open-system pingo origin^{1,17,18,21}. They are also sometimes located near stream junctions; a feature that has been taken to indicate a fluvial scour origin^{12,15}, but it is not unusual for open-system pingos to occupy such locations²³.

The deeper structures of active and fossil pingos have as yet been little explored. Nevertheless, with regard to the diapirism observed in the basal parts of the London depressions, it is noteworthy that the upward displacement of underlying strata into the base of pingo systems has been suggested in several instances²⁴⁻²⁶.

If the above proposals are valid, it will be possible to define more closely the areas of London where further anomalous depressions in the London Clay surface can be expected. Likewise, remnants of open-system pingos should be found in association with analogous artesian areas elsewhere in southern and central Britain.

We thank F. G. Berry, G. M. Johnson, L. M. Lake, J. R. Mackay, P. A. Mantz, J. Rose and W. H. Ward for helpful discussions.

Received 16 July 1979; accepted 6 February 1980.

1. Watson, E. *Phil. Trans. R. Soc. B* **280**, 183-198 (1977).
2. Pissart, A. Z. *Geomorphol.* **7**, 147-165 (1963).
3. Watson, E. *Geol. J.* **7**, 381-392 (1971).
4. Mitchell, G. F. *Nature* **230**, 43-44 (1971).
5. Sparks, B. W., Williams, R. B. G. and Bell, F. G. *Proc. R. Soc. A* **327**, 329-343 (1972).
6. Watson, E. and S. *Geogr. Ann.* **56A**, 213-225 (1974).
7. Burton, R. G. O. *East Midland Geographer* **6**, 230-240 (1974-77).
8. Institute of Geological Sciences *Quaternary Map of the United Kingdom* 1st edn (1977).
9. Bowen, D. Q. *Quaternary Geology* (Pergamon, Oxford, 1978).
10. Williams, R. B. G. in *The Periglacial Environment* (ed. Péwé, T. L.) 399-410 (McGill-Queen's University Press, Montreal, 1969).
11. Müller, F. *Meddr. Grønland* **153**, 1-127 (1959).
12. Berry, F. G. Q. *Jl Engng Geol.* **12**, 9-29 (1979).
13. Edmunds, F. H. *Mem. Geol. Surv. Summ. Prog. for 1930*, 18-24 (1931).
14. Higginbottom, I. E. & Fookes, P. G. Q. *Jl Engng Geol.* **3**, 85-117 (1970).
15. Zeuner, F. *The Pleistocene Period* (Hutchinson, London, 1959).
16. Rose, J., Turner, C., Coope, G. R. & Bryan, M. D. in *Timescales in Geomorphology* (eds Cullingford, R. A., Davidson, D. A. & Lewin, J.) 159-175 (Wiley, Chichester, 1980).
17. Mackay, J. R. *Permafrost: Proc. int. Conf.* **71**-76 (1966).
18. Brown, R. J. E. & Péwé, T. L. *Permafrost: N. Am. Contr. 2nd int. Conf.* **71**-100 (1973).
19. Wakeling, T. R. M. & Jennings, R. A. J. Q. *Jl Engng Geol.* **9**, 255-263 (1976).
20. Lucas, J. *Hydrogeological Survey* (Stanford, London, 1877).
21. Holmes, G. W., Hopkins, D. M. & Foster, H. L. *U.S. Geol. Surv. Bull.* **1241-H**, 1-40 (1968).
22. Williams, J. R. & van Everdingen, R. O. *Permafrost: N. Am. Contr. 2nd int. Conf.* **435**-446 (1973).
23. O'Brien, R. *Meddr. Grønland* **195**, 1-20 (1971).
24. Bik, M. J. J. *Biul. peryglac.* **19**, 85-130 (1969).
25. Flemal, R. C. *Proc. 24th int. Geol. Congr. Sect. 12*, 130-135 (1972).
26. Mackay, J. R. *Biul. peryglac.* **27**, 133-154 (1978).

Interaction between chemoreceptive modalities of odour and irritation

William S. Cain & Claire L. Murphy

John B. Pierce Foundation Laboratory and Department of Epidemiology and Public Health, Yale University, New Haven, Connecticut 06519

Inhaled vapours may stimulate both olfactory receptor cells and endings of the trigeminal nerve¹. The trigeminal nerve often contributes a pungent, irritating attribute to odours. Many odorants exhibit some pungency when dilute and most become pungent when concentrated². Because the trigeminal system commonly shares the chemosensory burden with olfaction, it is relevant to ask whether these anatomically distinct systems interact³⁻⁶. Two obscure psychophysical observations argue for an inhibitory influence of trigeminal over olfactory activity. Katz and Talbert observed that a vapour with both odour and pungency might lose odour at high concentrations, irritation masking odour⁷. The nineteenth century philosopher Alexander Bain, noting that concentrated carbon dioxide (carbonic acid) evokes pungency, remarked "if a current of carbonic acid accompanies an odour, the effect (odour) is arrested" (ref. 8). We have taken up Bain's forgotten observation and used carbon dioxide to endow otherwise benign concentrations of odorant with varying degrees of pungency. The experiments reported here reveal a strong mutual interaction between pungency and odour, occurring without attenuation even when irritant enters one nostril and odorant the other.

In the first experiment, eight persons judged the perceived magnitude of four concentrations of the fruity odorant *n*-amyl butyrate [0.7-12.2 parts per million (p.p.m.)], presented olfactometrically, four concentrations of carbon dioxide [10-50 parts per hundred (p.p.h.)], and all 16 combinations (mixtures) of the two. Subjects inhaled the stimuli through one nostril (flow rate: 4 l min⁻¹). They assigned numbers proportional to sensory magnitude⁹, judging a stimulus first for overall magnitude and then for either the odour component (half the trials) or the irritating component. Each subject made 240 judgments over five sessions.

In the second experiment, 10 subjects followed the same psychophysical procedure with dichorhnic 'mixtures' (4 l min⁻¹ flow rate into each nostril with equal numbers of right-left and left-right combinations of odorant and irritant). In such condi-

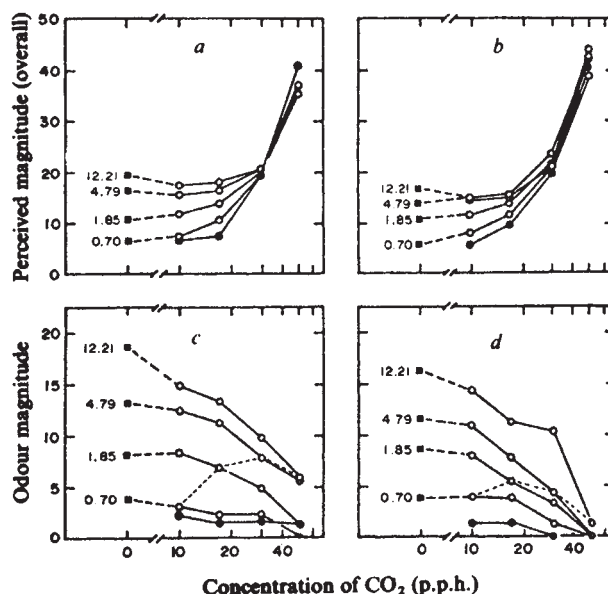


Fig. 1 *a*, Perceived magnitude (linear scale) plotted against concentration of carbon dioxide (logarithmic scale) for carbon dioxide presented alone (●), amyl butyrate presented alone (■) and mixtures of carbon dioxide and amyl butyrate (○). The concentration of amyl butyrate (p.p.m.) is indicated at the left in each case. Data points are medians taken across eight subjects. *b*, Same as *a*, but combinations of carbon dioxide and amyl butyrate were presented dichorhnicly, that is irritant (carbon dioxide) to one nostril and odorant (amyl butyrate) to the other. Data points are medians taken across 10 subjects. *c*, Perceived odour component (denoted odour magnitude) of amyl butyrate alone (■), carbon dioxide (●) and physical mixtures (○). The low, but non-zero judgments for the 'odour' of the odourless irritant carbon dioxide presumably reflect imperfect perceptual resolution between odour and irritation. The non-monotonic function formed by the thin dashes depicts how odour magnitude would change in a case where concentration of odorant and irritant changed jointly. *d*, Same as *c*, but dichorhnic mixtures.

tions, the nasal septum insulated odorant from irritant but the sensory 'image' was essentially indistinguishable from that produced by a physical mixture¹⁰.

If the olfactory and trigeminal systems did not interact, the families of functions in Figs 1*a*, *b* and 2*a*, *b* would comprise

parallel lines. The converging trend reveals inhibitory interaction. Figures 1c, d and 2c, d depict the pattern of the inhibition for each perceptual component. Just as the perceived magnitude of carbon dioxide alone increased sharply with concentration (Fig. 1a, b), so did its ability to inhibit any accompanying odour (Fig. 1c, d). Similarly, just as the perceived magnitude of amyl butyrate alone grew gradually with concentration (Fig. 2a, b), so did its ability to inhibit irritation (Fig. 2c, d).

Figures 1 and 2 reveal a fundamental similarity between the physical (monorhinal) and dichorhinal mixtures. Because it rules out interference between odourant and irritant at the mucosal surface, dichorhinal inhibition seems to establish the generality of the olfactory-trigeminal interaction beyond the stimuli used here. It also suggests that the interaction occurs in the brain.

The explanation of dichorhinal inhibition must not ignore possible bilateral peripheral phenomena, such as changes in nasal patency¹. Such changes could arise from reflex vasodilation, which could increase intranasal resistance and diminish accessibility of the olfactory mucosa¹. To account for inhibition of odour, however, this mechanism would require the trigeminal response to precede the olfactory response consistently, an unlikely outcome in view of previous results^{11,12}. The trigeminal system generally seems to react to chemical stimulation slowly. The third experiment supported this view in a comparison of reaction times to carbon dioxide alone and amyl butyrate alone (Fig. 3a). Six subjects each made 400 responses over five sessions. Both odour and irritation exhibited intensity-dependent latencies, but at any given level of perceived magnitude the reaction to amyl butyrate always preceded that to carbon dioxide. The reaction to carbon dioxide was too slow to support the possibility that it generally inhibited odour magnitude through reflex changes in nasal patency.

The fourth experiment gave temporal advantage to the irritant to investigate whether peripheral factors are involved in the interaction between odour and pungency. Thirteen subjects (each making 96 judgements over three sessions) assessed the magnitude of amyl butyrate after inhalation of either 30 p.p.h. carbon dioxide or pure air. Each subject sampled the adapting stimulus (carbon dioxide or air) for 2 s followed immediately, on

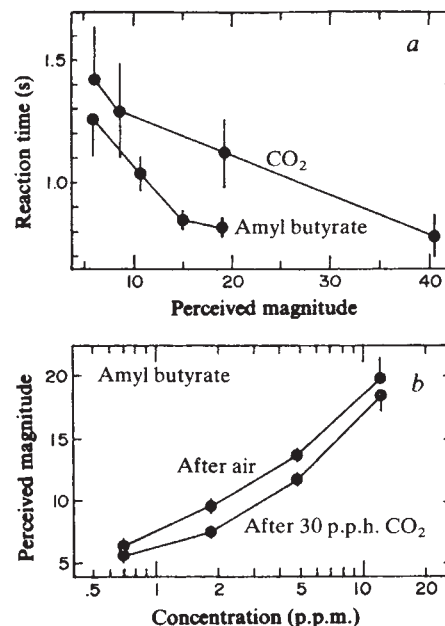


Fig. 3 a, Simple reaction time to four concentrations of amyl butyrate and to four concentrations of carbon dioxide plotted against the perceived magnitude of the stimuli. Data points are means (± 1 s.e.m.) taken across six subjects. b, Perceived magnitude of amyl butyrate smelt for 2 s immediately after 2-s inhalations of pure air (half the trials in each session, generally every other trial) or 30 p.p.h. carbon dioxide (other half of the trials). Data points are means (± 1 s.e.m.) taken across 13 subjects.

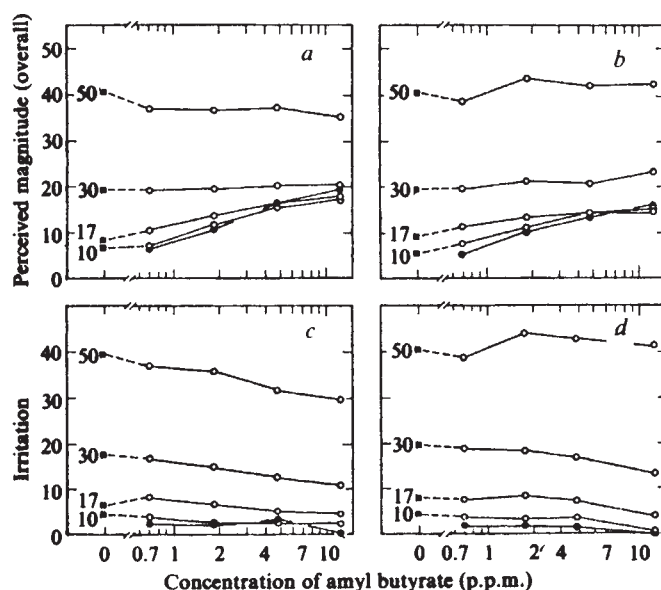


Fig. 2 a, Same psychophysical data as Fig. 1a, but plotted here against concentration of amyl butyrate. Symbols as follows: ●, amyl butyrate alone; ■, carbon dioxide alone; ○, physical mixtures. The concentrations of carbon dioxide (p.p.h.) are indicated at left for each. b, Same as a, but dichorhinal mixtures. c, Perceived irritating component of carbon dioxide alone (■), amyl butyrate alone (●) and physical mixtures (○). d, Same as c, but dichorhinal mixtures.

the same inhalation and through the same nostril, by amyl butyrate (2 s). In these conditions, carbon dioxide reduced odour intensity by about 12% (Fig. 3b). Simultaneous presentation had reduced odour intensity more than four times this amount (Fig. 2c, d). Accordingly, even in favourable temporal conditions, peripheral factors seem relatively impotent.

Because, if potent enough, an olfactory stimulus generally serves also as a trigeminal stimulus, olfactory-trigeminal interaction must be frequent. So far, the inhibitory interaction has revealed itself unambiguously only in the extreme case, such as where the inherent pungency of an 'odorant' outstrips odour magnitude, causing it to fall with increases in concentration^{7,12}. Figure 1c, d depicts how odour would first increase and then decrease with joint (co-varying) increases in the concentration of carbon dioxide and amyl butyrate. These non-monotonic functions could reflect how the odour of a single stimulus with both odour and pungency varies with concentration¹². Hints that such functions might emerge from central olfactory structures appear in observations that benign odorants will stimulate so-called background activity in the olfactory bulb, whereas 'sharp, unpleasant odors' will inhibit such activity^{13,14}. Walsh even noted a specific inhibitory effect of carbon dioxide on single units of the olfactory bulb (cats)¹⁵. These units also responded to the somesthetic (trigeminal) stimulation caused by inhalation of odourless air.

Thus, we have confirmed, quantified and extended a century-old observation that pungency can diminish odour. We show: (1) a continuum of inhibition, (2) mutuality, whereby odour inhibits irritation and vice versa, (3) apparent generality beyond particular olfactory or trigeminal stimuli, and (4) a central neural site of interaction. The dimensions of the interaction may depend on various factors. Irritants will undoubtedly vary in their affinity for olfactory receptors. Even carbon dioxide, although odourless, might conceivably cause some unspecified modulation of olfactory receptors. Manner of trigeminal activation may also have a role; perhaps trigeminally mediated warmth, cold or pressure can modulate odour. Finally, the timing of olfactory and trigeminal activation may be involved. Differences in speed

of activation from odorant to odorant and irritant to irritant, not necessarily revealed best by simple reaction time, may govern any contribution from peripheral factors. Irrespective of the mechanism of the interaction, our results suggest a need for special attention to odorous and irritating contaminants in the workplace. An irritating gas may mask the presence of useful olfactory signals (such as, warning agents) and an odorant may mask the possibly corrosive vapours (for example, inorganic acids) that often form the stimulus for the trigeminal nerve.

This research was supported by a grant from the NIEHS. We thank Mr Raymond Huey for technical assistance and Drs Robert C. Gesteland, John Kauer and M. M. Mozell for helpful criticism.

Received 4 September 1979; accepted 17 January 1980.

1. Tucker, D. in *Handbook of Sensory Physiology* Vol. VI, pt 1 (ed. Beidler, L. M.) 151–181 (Springer, Berlin, 1971).
2. Cain, W. S. *Ann. N.Y. Acad. Sci.* **237**, 28–34 (1974).
3. Stone, H., Carregal, E. J. & Williams, B. *Life Sci.* **5**, 2195–2201 (1966).
4. Stone, H., Williams, B. & Carregal, J. A. *Expl. Neurol.* **21**, 11–19 (1968).
5. Stone, H. in *Olfaction and Taste III* (ed. Pfaffmann, C.) 216–220 (Rockefeller University Press, New York, 1969).
6. Stone, H. & Rebert, C. S. *Brain Res.* **21**, 138–142 (1970).
7. Katz, S. H. & Talbert, E. J. *Tech. Pap.* 480 (Bureau of Mines, US Department of Commerce, 1930).
8. Bain, A. *The Senses and the Intellect*, 152 (Longmans Green, London, 1868).
9. Cain, W. S. & Moskowitz, H. R. in *Human Responses to Environmental Odors* (eds Turk, A., Johnston, J. W. Jr & Moulton, D. G.) 1–32 (Academic, New York, 1974).
10. Cain, W. S. *Nature* **268**, 50–52 (1977).
11. Tucker, D. in *Olfaction and Taste* (ed. Zotterman, Y.) 45–69 (Pergamon, London, 1963).
12. Cain, W. S. *Sensory Processes* **1**, 57–67 (1976).
13. Hughes, J. R. & Mazurkowsky, J. A. *Electroenceph. clin. Neurophysiol.* **14**, 646–653 (1962).
14. Sem-Jacobsen, C. W. *et al. Am. J. med. Sci.* **232**, 243–251 (1956).
15. Walsh, R. R. *Am. J. Physiol.* **186**, 255–257 (1956).

Illusory reversal of extrafoveally perceived displacement

D. M. MacKay

Department of Communication and Neuroscience, University of Keele, Keele, Staffordshire ST5 5BG, UK

It has long been known that the temporal characteristics of human peripheral vision differ markedly from those of foveal vision. Slowly moving peripheral stimuli, for example, can give rise to dramatically exaggerated estimates of their displacement^{1,2}. More recently, Thorson *et al.*^{3,4} have shown that where two spots are flashed in sequence to peripheral vision with an interflash interval of 50 ms, a sensation of movement can be induced even when the spatial separation of the spots is below the static acuity threshold. These observations fit with several others^{5,6}, suggesting that the visual system uses separate channels for signalling 'motion' as distinct from 'change of location', and that in some circumstances the integration of motion signals may make a dominant contribution to the perception of displacement. I now report a striking illusion which seems to reinforce and extend this conclusion.

This illusion is easily demonstrated and quantified using a double-beam cathode-ray oscilloscope with a non-persistent screen in a normally lit room. If one spot (or line) on the screen is deflected instantaneously through 1° or 2° and allowed to return more slowly to its original position, for example by alternate (1 s⁻¹) exponential-spike waveforms of 25 ms time constant and opposite polarity (Fig. 1a), then in foveal vision its motion is perceived more or less veridically. Viewed 3° or 4° extrafoveally, however, the appearance is surprisingly different. The displacement perceived is actually in the opposite direction to the real, the apparent motion being as shown in Fig. 1b. The illusion is readily quantified⁷ by imposing an RC-filtered step deflection of this form on the second cathode-ray tube beam, and adjusting its amplitude and time constant until the motion of the second

spot or line, when viewed foveally, seems to match that of the first (perceived extrafoveally). Detailed measurements are still in progress and will be reported elsewhere, but for photopic stimuli viewed at retinal eccentricities of a few degrees, the motion perceived is just as if the rapid displacements had not occurred. Furthermore, if fixation is changed so that both stimuli are viewed extrafoveally, their perceived motions still seem to match.

It thus seems that in these conditions only the slow phases of the image motion are signalled by the extrafoveal system to the centres that mediate motion perception. As the resting position of the spot (between spikes) remains physically unchanged (Fig. 1a), but is perceived as displaced alternately by an amount $\pm y$ (Fig. 1b), extrafoveally perceived location seems here to be computed entirely by integration from motion signals, rather as in an 'inertial navigation' system. (The rising phase of the spike, which takes only a few microseconds, is presumably too rapid to stimulate any physiological 'motion detectors' (ref. 8).)

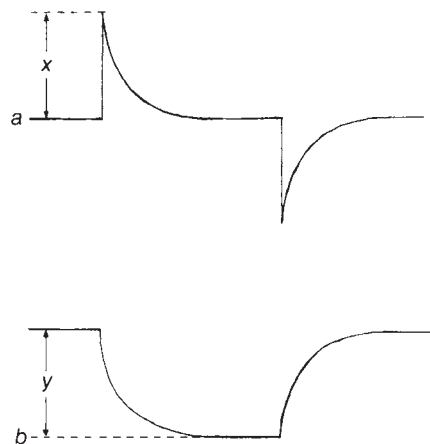


Fig. 1 a, Waveform of deflection of extrafoveally viewed spot (spike deflections $\sim 1^\circ$, 1 s⁻¹). b, Waveform of perceptually matching deflection of foveally viewed spot. For perceptual match, $y = x$ and time constants are equal.

Note that the displacement x (Fig. 1a) was here much larger than the static acuity threshold for the extrafoveal region stimulated. When a square wave of the same amplitude was substituted for a, the spot motion was clearly perceptible. The present illusion is thus distinct from, and in a sense complementary to, the Thorson effect^{3,4} described above.

Further experiments have shown that at scotopic light levels (with only the retinal rods active) the perceptual match between the foveal and extrafoveal stimuli of Fig. 1 breaks down. This suggests that the computation of extrafoveal displacement on the basis of motion signals is primarily a function of the cone system. In this connection it may be significant that it is normally only at scotopic light levels (when the cone system is inactive) that illusory instability of the perceived world is seen during saccadic eye movements. The findings reported here suggest that at photopic levels the processing of extrafoveal information may be undisturbed by the image displacements produced by normal saccadic exploratory movements.

Received 29 October 1979; accepted 21 January 1980.

1. Exner, S. *Pflügers Arch. ges. Physiol.* **11**, 403–432 (1875).
2. Kries, J. Von in *Handbuch der Physiologischen Optik* Vol. III, 3rd edn (ed. Southall, J. P. C., English edn) 275–276 (Optical Society of America, 1925).
3. Thorson, J., Lange, G. D. & Biederman-Thorson, M. *Science* **164**, 1087–1088 (1969).
4. Biederman-Thorson, M., Thorson, J. & Lange, G. D. *Vision Res.* **11**, 889–903 (1971).
5. MacKay, D. M. *Nature* **181**, 507–508 (1958).
6. MacKay, D. M. in *Handbook of Sensory Physiology* Vol. VII/3A (ed. Jung, R.) 307–331 (Springer, Heidelberg, 1973).
7. MacKay, D. M. *Proc. physiol. Soc.* September, 14P (1979).
8. Grüsser, O. -J. & Grüsser-Cornehlis, U. in *Handbook of Sensory Physiology* Vol. VII/3A (ed. Jung, R.) 333–429 (Springer, Heidelberg, 1973).

Young kittens can learn complex visual pattern discriminations

Frances Wilkinson

Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2

Peter C. Dodwell

Department of Psychology, Queen's University, Kingston, Ontario, Canada K7M 3N6

Kittens begin to display visually elicited responses shortly after eye opening. The pupillary reflex¹, optokinetic response^{2,3}, visual placing³⁻⁵ and avoidance of the deep side of the visual cliff^{3,5-7} all appear between the second and fifth weeks of life, and evidence of visual learning in homing behaviour has been described at the same early age⁸. However, an attempt to elicit pattern preferences in kittens of this age (usually considered the easiest way to demonstrate discriminatory behaviour in immature organisms⁹) was unsuccessful, evidently because such kittens show little spontaneous visual interest^{10,11}. This is surprising in view of electrophysiological findings which emphasise the importance of the fourth to sixth weeks in the development of cortical feature-analysing mechanisms¹²⁻¹⁴; perhaps the stimulus-seeking methodology of Dodwell *et al.*^{10,11} is inappropriate for the visual modality at such an early age. Recently, visual acuity measurements have been reported for kittens as young as 30 d of age using a contour versus no-contour discrimination on a modified Lashley jumping stand^{15,16}. With this technique we have now trained kittens of under 50 d of age to perform more complex pattern discriminations.

In the jumping stand, kittens have to jump from the lip of a short tunnel on to one of two stimulus surfaces, a drop of 15–20 cm. The surfaces are trapdoors, and an incorrect choice results in a fall of about 45 cm. Correct choices are rewarded with food and petting. An average kitten is able to perform this task adequately at 28–30 d of age, thus setting a lower age limit to its use. To accustom the animals to the jumping stand and to obtain an initial estimate of their learning ability, all kittens are pretrained on a white/black discrimination (white positive). The 20 kittens in our experiment all exhibited either an immediate preference for white, or learning within 80 trials. Accustoming the animal to jumping, black/white training, and gradually increasing the jumping height to a maximum of 15–20 cm, takes 5–10 d, thus placing the onset of pattern discrimination training between the 35th and 40th day of life.

Different groups of kittens of this age were trained to discriminate each of the three pattern pairs shown in Fig. 1. These stimuli constitute the three basic orthogonal orbit-pairs of Hoffman's Lie transformation group theory of pattern perception (ref. 17 and P.C.D., unpublished) and were chosen for that reason; the theory predicts that they will be the easiest patterns to discriminate, particularly for an immature animal¹⁷. However, patterns similar to pairs *a* and *b* have been widely used in the study of visual perception (refs 18–20 and P.C.D., unpublished) pair *b* in particular having been generally regarded as the simplest possible pattern discrimination because its resolution involves only orientation detection²⁰. Pair *c* (families of rectangular hyperbolae) have not been used before. The stimuli were photographic enlargements of these patterns mounted on 32-cm² white cards and plasticised. The black contours were 1.5 cm wide, thus subtending visual angles of slightly more than 4° of arc when viewed from a height of 20 cm. This should be well within the visible range for kittens at 35 d of age¹⁵. Each stimulus pair was matched as closely as possible for overall luminous flux.

Kittens were given 40 training trials per day until the criterion of 27 of 30 consecutive trials correct was met. The number of trials (excluding the final 30) and errors to criterion for the three groups are shown in Table 1, with the age at which criterion was reached. Clearly, kittens are able to master each of these problems very early in life. For comparison, we have included data for problem *a* from seven older kittens (mean age at criterion 80 d), some of which had previously learned other problems and some of which were experimentally naive. Both young and older kittens showed similar performance, so apparently the young group made use of visual pattern information in a normal manner throughout the training period, and learning was not impaired by visual or motor immaturity. This is probably true also of problems *b* and *c*, but we do not yet have sufficient data to confirm the point. The results in Table 1 suggest that for very young kittens the problems increase in difficulty from *a* to *c*. We are now investigating whether or not this is an age-dependent phenomenon.

Solution of each discrimination problem was followed by a series of transfer tests. These were designed to rule out the involvement of various extraneous cues and to determine to which aspects of the stimulus configuration the animals were responding. Transfer tests were typically administered by a different investigator from the one who administered the training trials; in no case did performance deteriorate as a result of this change. On the basis of the perfect or near-perfect transfer scores obtained, such factors as olfactory cues, auditory and visual cues associated with locking the trapdoors, and visual cues specific to the particular pair of stimulus cards used in training could be ruled out. A full description of the methodology and results of these tests will be published elsewhere.

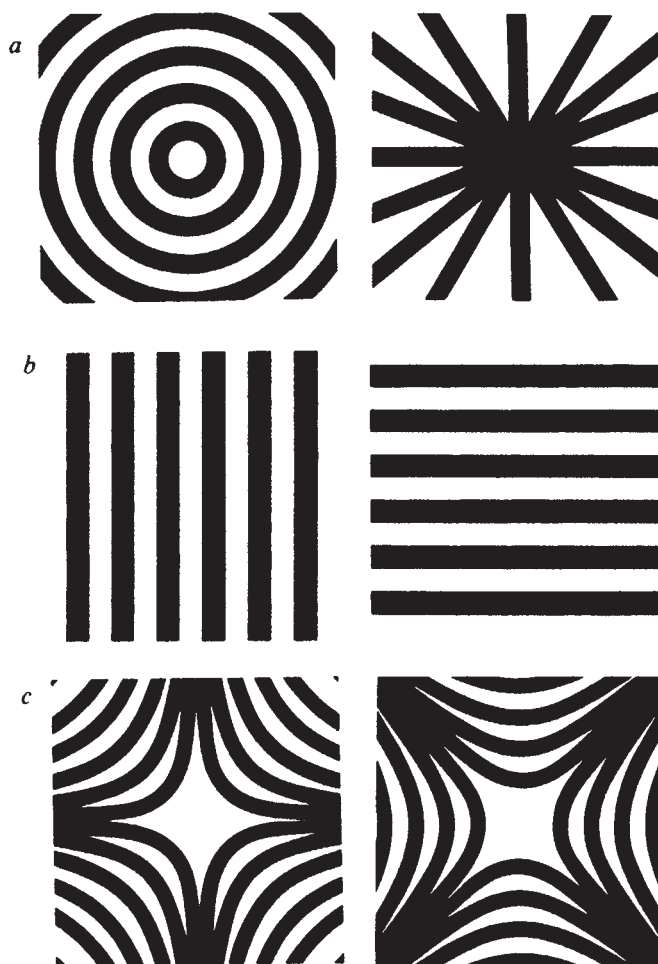


Fig. 1 The three pattern pairs derived from Hoffman's Lie transformation group theory.

Table 1 Summary of learning scores

Problem	No. of kittens	Trials to criterion	Errors to criterion	Age at onset of training (d)	Age at criterion (d)
a	9	115 (40–218)	53 (19–99)	38 (35–40)	43 (38–47)
b	4	271 (97–435)	113 (49–225)	38 (35–40)	48 (38–58)
c	4	327 (163–562)	138 (79–239)	38 (33–43)	48 (46–51)
a (late)	7	119 (42–187)	53 (21–89)	76 (45–120)	80 (48–123)

Results given are the mean; the range is given in parentheses. Group a (late) consisted of the four kittens from problem b group, one kitten from problem c group, and two experimentally naive kittens.

For problems a and c the most crucial test was the contrast-reversal test, in which the stimuli were negative photographic enlargements of the same original patterns. The contrast relationships within each figure were thus reversed, as was the direction of any overall luminance difference between figures. This test was given to 14 of the 16 animals trained on problem a and three of the four kittens trained on problem c. Contrast-reversal test trials were intermixed with training trials on the original stimulus pair. Food reward followed all test trials regardless of the animal's response whereas, on training trials, the reinforcement contingencies described above were used. The range of transfer scores fell between 7/10 and 10/10 correct, with a mean of 9.0. This constitutes fairly strong evidence that the kittens were not solving the problems on the basis of overall luminous flux or local brightness cues, thus suggesting that they coded the features of the patterns as a whole, and independently of contrast.

We have thus demonstrated that kittens between 30 and 50 d of age can be taught arbitrary associations between certain visual patterns and positive and negative rewards. This seems almost paradoxical in view of the earlier finding that kittens show little, if any, spontaneous interest in visual stimulation at this age¹⁰. The apparent anomaly can probably be accounted for in terms of the extrinsic rewards associated with our present training procedure. Nevertheless, the findings reported here open the way for investigating the ontogeny of visual pattern analysis under much tighter experimental control than is provided by the naturalistic situations in which visual learning has previously been described in very young kittens⁸. Our initial results indicate that at the peak of the so-called 'critical period', when neurones of the visual cortex are most sensitive to environmental modification^{21,22}, kittens are able to extract and use visual pattern information in a rather sophisticated manner, albeit from patterns which, from a certain theoretical point of view, are the simplest ones available.

This work was supported by grants from NSERC of Canada to F.W. (A7076) and P.C.D. (A0A44). We thank Kerry Sue Hare and Linda Sorensen for technical assistance.

Received 15 October 1979; accepted 23 January 1980.

- Windle, W. F. *J. comp. Neurol.* **50**, 479–503 (1930).
- Warkentin, J. & Smith, K. U. *J. Genet. Psychol.* **50**, 371–399 (1937).
- van Hof-van Duin, J. *Brain Res.* **104**, 233–241 (1976).
- Riesen, A. H. *J. Nervous Ment. Dis.* **132**, 21–25 (1961).
- Norton, T. T. *J. Neurophysiol.* **37**, 674–690 (1974).
- Walk, R. D. & Gibson, E. J. *Psychol. Monogr.* **75**, (1961).
- Karmel, B. Z., Miller, P. H., Dettweiler, L. & Anderson, G. *Dev. Psychobiol.* **3**, 79–90 (1970).
- Rosenblatt, J. S., Turkewitz, G. & Schneirla, T. C. *Trans. N.Y. Acad. Sci.* **31**, 231–250 (1964).
- Fantz, R. L., Fagan, J. F. & Miranda, S. B. in *Infant Perception* Vol. 1 (eds Cohen, L. B. & Salapatek, P.) (Academic, New York, 1975).
- Dodwell, P. C., Timney, B. N. & Emerson, V. F. *Nature* **260**, 777–778 (1976).
- Timney, B. N., Emerson, V. F. & Dodwell, P. C. *Q. J. exp. Psychol.* **31**, 63–81 (1979).
- Pettigrew, J. D. *J. Physiol., Lond.* **237**, 49–74 (1974).
- Blakemore, C. & Van Sluysers, R. C. *J. Physiol., Lond.* **248**, 663–716 (1975).
- Buisseret, P. & Imbert, M. *J. Physiol., Lond.* **255**, 511–525 (1976).
- Mitchell, D. E., Giffin, F., Wilkinson, F., Anderson, P. & Smith, M. L. *Vision Res.* **16**, 363–366 (1976).

- Mitchell, D. E., Giffin, F. & Timney, B. N. *Perception* **6**, 181–193 (1977).
- Hoffman, W. C. *J. math. Psychol.* **3**, 65–98 (1966); **4**, 348–349 (1966).
- Fantz, R. L. *Ann. N.Y. Acad. Sci.* **118**, 793–814 (1965).
- MacKay, D. M. *Nature* **180**, 849–850 (1957).
- Dodwell, P. C. *Visual Pattern Recognition* (Holt, Rinehart and Winston, New York, 1970).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **206**, 419–436 (1970).
- Blakemore, C. & Van Sluysers, R. C. *J. Physiol., Lond.* **237**, 195–216 (1974).

A demonstration of navigation by small rodents using an orientation cage

Janice G. Mather & R. Robin Baker

Department of Zoology, University of Manchester, Manchester M13 9PL, UK

The ability to 'home' after displacement has been demonstrated for several species of small mammals¹. Whether such homing is accomplished by random scatter², familiarity with a large area^{3,4} or some kind of navigation mechanism⁵ is still unknown. The ability to navigate (to determine the direction that will lead the animal to the required destination⁶) has been shown, for example, by releasing displaced rodents on snow and analysing their tracks⁶ and by analysing for homeward bias the directions from the release point of traps in which displaced individuals were subsequently recaptured¹. Such techniques are limited in application. Moreover, the latter technique is susceptible to differential avoidance of traps by animals travelling in different directions³. It is potentially more useful to retain displaced individuals in an orientation cage and to examine their movements within the cage for evidence of orientation in the homeward direction. However, previous attempts using various types of orientation cages have yielded only negative results^{8–10}, except, surprisingly, when the displaced animals could not see their natural surroundings. We present the first positive evidence that small rodents will navigate while retained in an orientation cage, even when able to see their surroundings. Indeed, we show that these visual cues are important in navigation and suggest why orientation cages have previously led to negative results.

Our main experimental species was the European wood-mouse, *Apodemus sylvaticus* although the yellow-necked mouse, *A. flavicollis*, and bank vole, *Clethrionomys glareolus*, were occasionally tested when available. In all, 19 individuals were used (*A. sylvaticus*, ten males and four females; *A. flavicollis*, one male; *C. glareolus*, two males and two females) in a total of 116 4-minute test periods on 72 displacement occasions. Although mainly nocturnal, *Apodemus* spp. are known also to be active during the day, particularly juveniles (unpublished observations and refs. 11 and 12). It was presumed, therefore, that if they had navigational ability, these species, like the more diurnal *C. glareolus*, ought to be able to make use of cues available during the day as well as the night. It was also anticipated, and borne out by this work, that a nocturnal animal would show a strong homing response during daylight.

All experiments were conducted between 10.00 and 17.30 h GMT between 11 and 25 July 1979 at Woodchester Park Field Centre, Gloucestershire. Animals were caught in Longworth traps, either in a meadow (12 traps in a 110-m line) or in woodland (96 traps in a 70 m × 110 m grid described by Yalden¹³). The woodland was situated above the meadow on the same north-west facing slope of an enclosed V-profiled valley. The orientation cage designed for this work and the experimental procedure are described in Fig. 1 legend. After testing, the animal was either returned to the point of capture or released from the final displacement site and the vanishing point direction (the direction of last observed position from the test site) recorded. Vanishing point direction (up to 30 m from the test site) was often very different from the initial direction of movement after release, which, as found by other authors⁷, was usually aimed at the nearest cover.

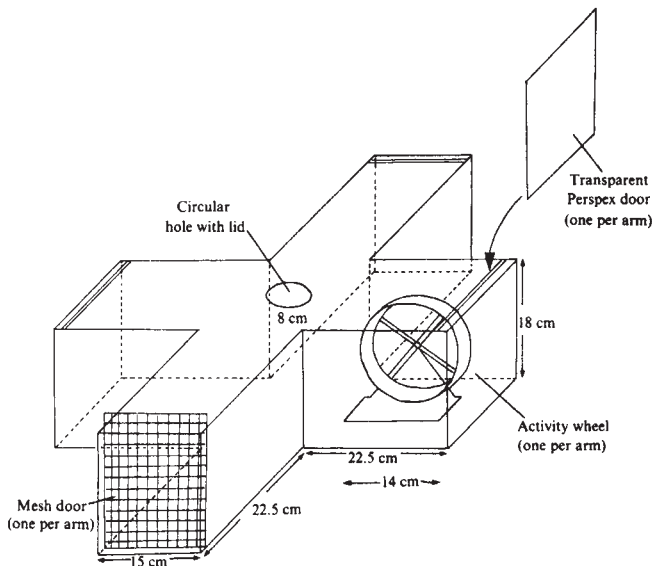


Fig. 1 The orientation cage was a cross-shaped Perspex chamber with a black opaque floor and transparent sides and top. A mesh door at the end of each arm of the cross could be raised to allow the animal to exit. An activity wheel was positioned in each arm as part of a separate experiment but which incidentally (and importantly) provided shelter. After capture, each individual was identified and weighed, then placed in the apparatus through a circular hole in the centre of the lid. An opaque black polythene sheet was used to cover the apparatus during transport to each of the displacement sites where an area of level ground was selected. The apparatus was then positioned so that the four arms were directed towards magnetic north, south, east and west. The polythene sheet was then removed and the observers withdrew to a vantage point ~5 m away. The experimental animal was allowed to visit three of the four arms before recording began. Total time spent in each of the four arms was then recorded, but periods of grooming, which were often long, were excluded. The danger of subjectivity thus introduced was reduced by the fact that two observers were always present, each with a stopwatch. In practice, the beginning and end of grooming periods were clear-cut and exclusion was considered preferable to lumping the behaviour with periods of movement and environmental monitoring. Recording ceased after 4 min (excluding grooming time). Whenever possible, this experiment was repeated with the same individual at four displacement locations, each varying in distance and compass direction from the central point of capture. At each displacement site, the direction and distance of the trap, observers and nearest vegetation/cover were recorded and also weather conditions and time of day. The influence of smell of previous occupant was controlled by: (1) maintaining the same arm of the apparatus in a constant compass direction during each day's experiments; (2) varying the compass sector from 'home' (see Fig. 2) of the first test location for successive individuals; and (3) testing each individual at locations in more than one compass sector from home. Once navigation in the equipment had been demonstrated, possible visual and olfactory mechanisms were examined. At each displacement site various experiments were conducted involving the use of a visual barrier (a cardboard screen, 60 cm high and 35 cm radius, with a small hole 1 cm in diameter for observation) which surrounded the apparatus, and a partial olfactory barrier (perspex doors which were slotted in alongside the mesh doors).

Statistical analysis was carried out in two stages as described by Batschelet^{14,15}. Time spent in each of the four arms of the orientation cage was used in a first-order analysis to calculate the mean vector (r, θ) for each 4-minute test (Fig. 1). The mean angles were then subjected to second-order analysis¹⁵. Non-uniformity was examined by Rayleigh's Z -test¹⁴. Stephen's exact test and Wheeler and Watson's parametric two-sample test¹⁴ were also used, as indicated in Table 1 and Figs 2, 3.

In displacement experiments, the interpretation of navigational ability is very much dependent on the experimenters making the correct guess concerning the animal's intended destination. We had to presume that the animals were residents and that the point of capture was located centrally within the individual's home range, to which it would want to return after

being displaced. Hence, point of capture was taken to be the location of home towards which the animal would orientate while retained within the apparatus, and statistics are presented with trap direction designated as 0° . However, this is a particularly stringent criterion as the directions of the trap and intended destination are unlikely to be identical. Nevertheless, our results (Table 1) show that on average the mean orientation in the apparatus and the direction of movement after release are not significantly different and that vanishing point direction is not significantly different from the direction of the trap. We conclude that the mean orientation in our cage reflects the direction in which the animal would travel (after seeking cover) if unrestrained and that this direction is towards home.

The data were analysed with respect to a number of potentially confounding variables and the results are summarised in Table 1. The number of occasions an individual was caught in a Longworth trap does not influence performance. In contrast, the number of previous uses in the test apparatus does. Individuals used for the first and second times show a preferred orientation near to 'home' direction. However, for some reason as yet unknown, individuals being used on their third occasion showed uniform orientation.

The preferred orientation was not influenced by the position of the observer or the nearest vegetation cover, but was affected by weather conditions. At wind speeds less than force 3 (Beaufort Scale) orientation is most marked when the Sun is visible but is non-uniform even when not. However, winds of Force 3 or greater disrupt orientation with respect to home,

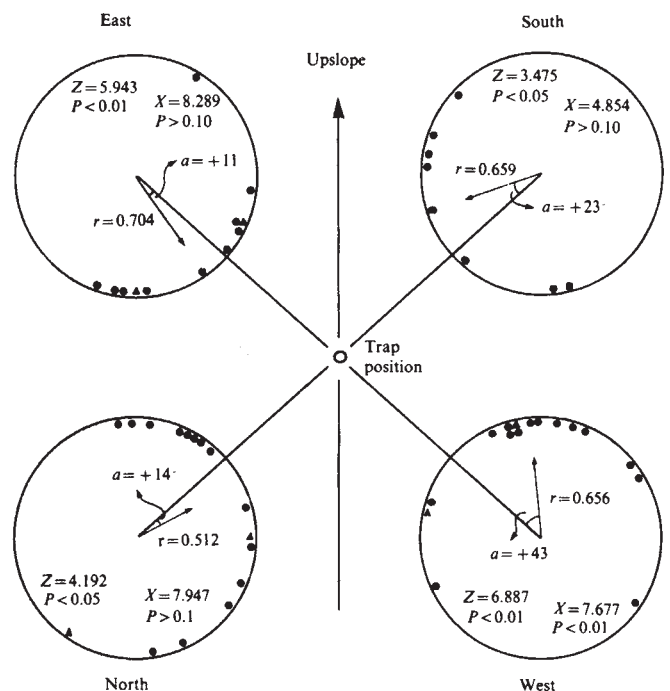


Fig. 2 Effect of displacement direction on mean orientation was shown by lumping data for displacements in four compass sectors around point of capture (trap position). All displacements 6–80 m from the trap are presented (excluding tests: (1) in wind speed of Force 3 or greater; (2) involving individuals being used on their third or more occasion; and (3) with a visual barrier). Each point (●, *Apodemus sylvaticus*; ■, *A. flavicollis*; ▲, *Clethrionomys glareolus*) represents the mean angle (θ , first-order analysis) for an individual during a 4-minute test period. Mean angle is expressed as positive or negative (+ or -) deviation from direction of trap of capture. All second order mean vectors (arrows) are significantly non-uniform. Mean vectors from N, E and S are not significantly different from trap direction (Stephens' exact test). From the W sector, the mean vector, although towards the trap, is significantly different from the trap direction. However, the inclusion of short-distance displacements (6–20 m) imposes a positive bias to the mean vector (see Fig. 3), an effect that possibly may have been compounded by some unslope orientation from this sector. We conclude that our animals were showing navigation in the orientation cage in the conditions and at the distances tested.

Table 1 Effect of various factors on orientation performance in the test apparatus: summary of second-order analyses

Variable	No. of tests (n)	Mean angle (θ°)	Mean vector (r)	Z	X	F	Exclusions*
<i>Home = 0°</i>							
Comparison of vanishing point direction (VP) with mean direction (MO) in test apparatus							
VP	13	+28	0.511	3.398†	5.868}	0.043	A, E
MO	13	+21	0.510	3.381†	6.190}		
No. of times caught:							
1 or 2	29	+19	0.531	8.172‡			B, C, D, E, F
3-16	20	+31	0.673	9.056‡			
Performance on successive trials:							
first use	36	+24	0.579	12.077‡			B, C, D, F
second use	13	+25	0.604	4.740‡			
third use	13	-74	0.220	0.627			
Effect of weather:							B, C, E, F
Wind speed < force 3							
with Sun	32	+25	0.631	12.737‡			
Sun obscured	7	+44	0.758	4.024†			
Wind speed > force 3	12	+49	0.183	0.403			
with Sun	5	-130	0.085	0.036			
Sun obscured	7	+49	0.375	0.983			
Effect of habitat:							B, C, E, F
open meadow	24	+21	0.567	7.715‡			
woodland	28	+26	0.641	11.488‡			
Sex							B, C, E, F
male	26	+29	0.560	8.147‡			
female	26	+17	0.645	10.822‡			
Reduced cues due to							B, C, D, E
opaque screen	8	+15	0.274	0.600			
Perspex doors	11	-3	0.572	6.291‡			
Observer direction = 0°	68	172	0.075	0.386			D, E, F
Nearest vegetation = 0°							B, C, E, F
trap distance							
6-80 m	46	49	0.176	1.421			
>80 m	5	89	0.289	0.419			
Upwind = 0°	14	78	0.475	3.164†			F, G

* A, animals returned to trapping site for release; B, test <6 m from trap; C, tests >80 m from trap; D, tests at wind speeds force 3; E, individuals on third time trial or more; F, tests with opaque screen; G, tests at wind speeds <force 3.

† $P < 0.05$.

‡ $P < 0.01$.

Z, Rayleigh's test; X, Stephens' exact test; F, Wheeler and Watson's 2-sample test.

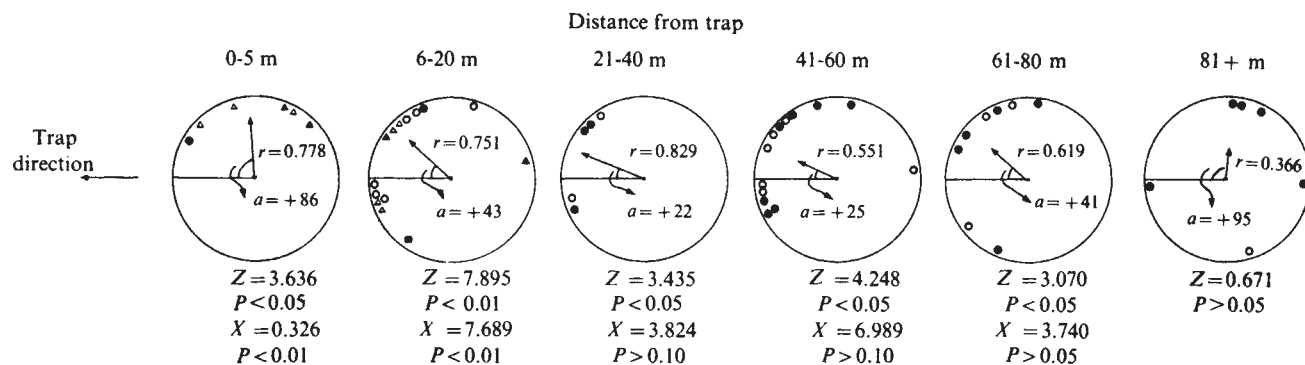


Fig. 3 The accuracy of orientation was influenced by displacement distance. Displacements in all directions from the trap are shown, excluding tests influenced by wind and previous usage as described in Fig. 2. Diagrammatic conventions are as in Fig. 2 except that open symbols denote females and closed symbols males. All second-order mean vectors (arrows) for distances less than 80 m are significantly non-uniform. At short distances (<20 m), mean vectors are significantly different from trap direction (Stephens' exact test), probably reflecting the difference between the directions of the trap and intended destination, although the meaning of the clear positive deviation is unknown. At greater distances a positive bias is still evident, though not significantly different from trap direction. Beyond 80 m, orientation is not significantly non-uniform, but sample size is small.

whether the Sun is visible or not. Instead, at these wind speeds the animals orientated at right angles to upwind, thereby spending more time in the most sheltered arm of the apparatus.

Final demonstration of the occurrence of navigation towards home in the apparatus was achieved by examining mean orientation from four different compass sectors relative to trap position (Fig. 2). Navigation has been demonstrated for distances up to 80 m from point of capture (Fig. 3) which is nearly three times the distance of half the home range length (0.5×56.6 m) of male *A. sylvaticus* at Woodchester in summer¹⁶. However, 80 m is far less than the distance from which homing has been demonstrated for *Apodemus* spp. elsewhere (59% return from 250 m,

17% return from 750 m etc.)¹. This could imply that navigation is not used at these longer distances and that homing is achieved by other mechanisms such as random search (see Bovet¹ for discussion). However, our experimental animals were limited to cues available at the particular displacement site and had to reach a decision concerning home direction within a few minutes. Further experiments at longer distances are planned with the animals being allowed more opportunity to acquire and process information as to their location.

Other factors were examined (Table 1) but were found to have no significant effect on the direction of orientation. These factors included whether the displacement site was in open meadow or

woodland and time of day (the latter is not shown in Table 1). Nor were any overall differences found between males and females (see also refs 4 and 17). However, when displacement distance was considered, there was an indication that orientation was most accurate for females at distances less than 40 m but for males at distances between 40 and 60 m (Fig. 3).

An opaque screen surrounding the apparatus disrupted orientation (Table 1), supporting the suggestion that woodmice use visual cues for navigation^{18,19}. The addition of Perspex doors (in the absence of a screen) (Fig. 1) did not disrupt orientation (Table 1). However, all 11 individuals thus tested showed a decreased bias (that is, first-order r value) towards the mean angle compared with the bias when the Perspex doors were removed. This was due to an increase in the amount of movement in the apparatus when the doors were in position, all arms being visited more frequently and evenly. During this movement, the animals often climbed the bars to sniff at any gaps still remaining. Vision thus seemed to be the major sense involved in navigation but additional information may have been sought from olfaction.

Our experiments have shown that in certain conditions small rodents will orientate toward home while in an orientation cage and will do so at distances up to at least three times the radius of their home range. Visual and perhaps olfactory senses both seem to be involved. We suggest that our experiments produced

positive results because, unlike previous orientation cages, ours offered the animal shelter from the wind and the opportunity to take cover (under the running wheels) should it feel threatened. As a result, the animals orientated toward home rather than toward the nearest cover, the main cause of negative results in previous experiments.

Received 16 August 1979; accepted 17 January 1980.

1. Bovet, J. in *Animal Migration, Navigation and Homing* (eds Schmidt-Koenig, K. & Keeton, W. T.) 405–412 (Springer, Heidelberg, 1978).
2. Murie, M. J. *Mammal* **44**, 338–349 (1963).
3. Robinson, W. L. & Falls, J. D. *Am. Midl. Nat.* **73**, 188–224 (1965).
4. Fidler, G. F. *J. Mammal* **47**, 53–58 (1966).
5. Bovet, J. Z. *Tierpsychol.* **19**, 472–88 (1962).
6. Bovet, J. Z. *Tierpsychol.* **28**, 211–216 (1971).
7. Baker, R. R. *The Evolutionary Ecology of Animal Migration* (Hodder & Stoughton, London, 1978).
8. Lindenlaub, E. Z. *Tierpsychol.* **12**, 452–458 (1955); **17**, 578 (1960).
9. Bovet, J. Z. *Tierpsychol.* **17**, 728–755 (1960).
10. Fidler, G. F. *Can. J. Zool.* **45**, 261–268 (1967).
11. Elton, C., Ford, E. B., Baker, J. R. & Gardner, A. D. *Proc. zool. Soc. Lond.* 657–721 (1931).
12. Miller, R. S. *Proc. zool. Soc. Lond.* **125**, 505–519 (1954).
13. Yalden, D. W. *J. Zool., Lond.* **164**, 244–250 (1971).
14. Batschelet, E. in *Animal Migration, Navigation and Homing* (eds Schmidt-Koenig, K. & Keeton, W. T.) 3–24 (Springer, Heidelberg, 1978).
15. Batschelet, E. *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms* (Am. Inst. Biol. Sci., Washington DC, 1965).
16. Montgomery, W. I. thesis, Univ. Manchester (1977).
17. Fidler, G. F. *Am. Midl. Nat.* **68**, 357–368 (1962).
18. Cooke, J. A. & Terman, C. R. *J. Mammal* **58**, 58–66 (1977).
19. Lüters, W. & Birukow, G. *Naturwissenschaften* **50**, 737–738 (1963).

Oestrogens regulate divergent effects of prolactin in the ovary

Johannes D. Veldhuis & James M. Hammond

Department of Medicine, Division of Endocrinology, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033

Prolactin exerts well established stimulatory effects both *in vitro* and *in vivo* in the rat ovary¹. However, in the ovary of higher mammals, prolactin action is not well characterised, and may differ from that observed in rodents². For example, in human clinical conditions of physiological or pathological prolactin excess, ovarian function is depressed³, and in isolated, human granulosa cells *in vitro*, prolactin seems to inhibit progesterone production⁴. To examine direct prolactin effects in the higher mammalian ovary, we have used an *in vitro* porcine granulosa-cell model⁵. In this system, prolactin action is bipotential, depending critically on the degree of granulosa-cell differentiation attained *in vivo*. Prolactin suppresses steroid production by cultured granulosa cells isolated from small (1–2 mm), immature follicles, but stimulates progesterone secretion by granulosa cells collected from large (>6 mm) mature follicles⁵. The present studies show that oestrogens may play an important part in regulating these divergent actions of prolactin in the ovary.

To investigate potential modulation of prolactin action by oestrogens, we maintained granulosa cells collected from immature porcine follicles in monolayer culture for 10–12 d. As previously observed, prolactin alone produces inhibitory effects (Fig. 1a). However, the continuous administration of oestrogens to granulosa cells produces a biphasic response and a striking oestrogen–prolactin interaction that is time dependent. Early in culture (days 0–2) oestradiol alone exerts dose-dependent inhibition of progesterone production. Oestradiol decreases the total mass of progesterone contained in cells and medium (data not shown), indicating a reduction in actual steroid synthesis rather than in steroid secretion alone. The range of inhibitory concentrations of oestrogens used *in vitro* encompassed that observed in porcine follicular fluid comprising the *in vivo* micro-milieu of immature to increasingly mature granulosa cells⁶. When prolactin is added to submaximally inhibiting concen-

trations of oestrogen during this initial culture period, progesterone secretion declines further. Thus, both oestrogen and prolactin exert inhibitory effects early in culture and these effects are additive at submaximal steroid concentrations.

After 48 h of continued oestrogen exposure, cultured granulosa cells exhibit a reversal in responsiveness to oestrogen effects, and show enhanced progesterone secretion. From day 2 onward in culture, oestradiol stimulates steroidogenesis, and, more strikingly, produces a 'switch' in prolactin action from inhibitory to stimulatory (Fig. 1b). The interaction of oestradiol and prolactin is synergistic in enhancing progesterone production on all days throughout culture after day 2 ($F > 20$, $P < 0.01$ by 2-way analysis of variance). The oestrogen effects exhibit steroid specificity. The stimulatory actions of oestrone ($1 \mu\text{g ml}^{-1}$) and 17β -oestradiol ($1 \mu\text{g ml}^{-1}$) were quantitatively similar, at 7.4 ± 0.7 and 8.7 ± 0.6 ng progesterone per 10^6 cells per 48 h, respectively, with control cultures 2.8 ± 0.3 (mean $\pm$ s.e.m., $n = 4$, $P < 0.01$ by one-way analysis of variance). However, the less potent oestrogen, oestriol ($1 \mu\text{g ml}^{-1}$), exerted no significant effect (2.1 ± 0.25), and the anti-oestrogen, nafoxidine hydrochloride (10^{-7} M), produced slight but insignificant suppression of steroid secretion (1.6 ± 0.2). Although oestrogens are mitogenic for granulosa cells *in vivo* and *in vitro*⁷, oestrogenic effects cannot be accounted for by simple alteration in cell number. 17β -oestradiol ($1 \mu\text{g ml}^{-1}$) increased the number of cells per culture by only $28 \pm 9\%$ above control ($n = 18$ experiments). Thus, when progesterone accumulation is corrected for cell density, both the acute inhibitory and the delayed stimulatory changes in progesterone secretion persist at various times in culture (Fig. 2).

In addition to modifying granulosa-cell responsiveness to prolactin, oestrogens seem to exert important intrinsic effects on steroid secretion by the ovary. *In vivo* and *in vitro*, oestrogens in conjunction with gonadotropins facilitate follicular maturation and steroidogenesis in several species^{7–9}. However, inhibitory effects of oestrogen on granulosa cell progesterone secretion have also been described^{10–14}. The present study demonstrates that the action of oestrogen may be bipotential, and critically dependent on the duration of hormone exposure. Acute oestrogen administration produces inhibitory effects, concordant with observations by Haney *et al.* in porcine granulosa cells¹⁰, and with work by other investigators in luteal tissue of the cow¹¹, ewe¹², monkey¹⁴ and human¹⁴. However, our findings also reveal delayed, but sustained stimulatory effects of oestrogen. A similar response to chronic oestrogen exposure in

replicating cultures has been reported by Bernard in rat¹⁵ and by Goldenberg *et al.* in pig granulosa cells *in vitro*¹⁶, but was not observed by Thanki and Channing when cultured porcine granulosa cells were maintained in serum-restricted conditions of stationary growth¹⁷. These observations suggest that factors related to cell proliferation must also be considered in the expression of oestrogen action in the ovary.

The directional reversal of prolactin action induced by oestrogens *in vitro* is similar to that which occurs spontaneously during maturation of porcine ovarian follicles *in vivo*^{5,18}. Granulosa cells of the pig may differ in this regard from those of the human (in the latter, prolactin seems to suppress progesterone secretion regardless of follicle size)⁴. This difference may reflect species characteristics, or exemplify the requirement to select mature, 'pre-ovulatory' follicles. Such follicles are exposed to high endogenous oestrogen concentrations *in vivo*, and are required in our system to demonstrate stimulatory effects of prolactin *in vitro*. Because intrafollicular oestrogen

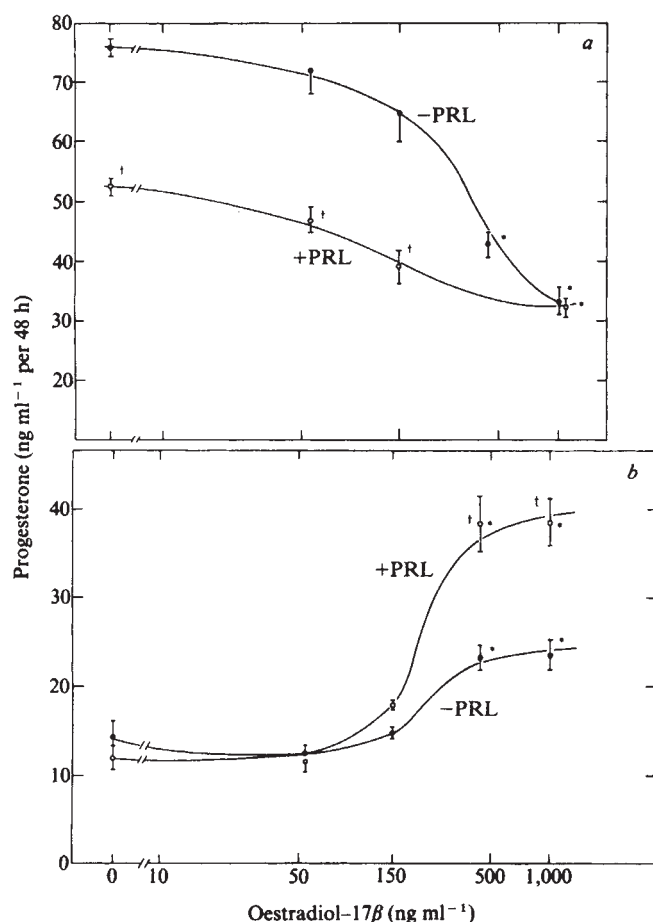


Fig. 1 *a*, Day 2, Acute oestrogen administration to cultured porcine granulosa cells suppressed progesterone production in a dose-dependent manner. Prolactin (PRL) (100 ng ml⁻¹) further inhibited progesterone accumulation in the presence of submaximal oestradiol concentrations. *b*, Day 4, Continuous oestrogen exposure enhanced progesterone secretion and reversed the inhibitory action of prolactin. The interaction of oestrogen and prolactin was synergistic in augmenting steroid accumulation ($F > 20$, $P < 0.01$ by 2-way analysis of variance). Granulosa cells were aspirated from 1–2-mm follicles of porcine ovaries and maintained for 2–10 d in monolayer culture (medium 199, 10% fetal calf serum, bicarbonate buffer and antibiotics⁵). Quadruplicate cultures were treated with increasing concentrations of 17β-oestradiol, in the presence or absence of prolactin (100 ng ml⁻¹, NIH-oPRL-S12). Spent medium was removed at 48-h intervals for the subsequent determination of progesterone by radioimmunoassay²³. Data are mean ± s.e.m. ($n = 4$). * $P < 0.01$ compared with basal (no oestrogen, no prolactin); † $P < 0.01$ compared with oestrogen treatment without prolactin.

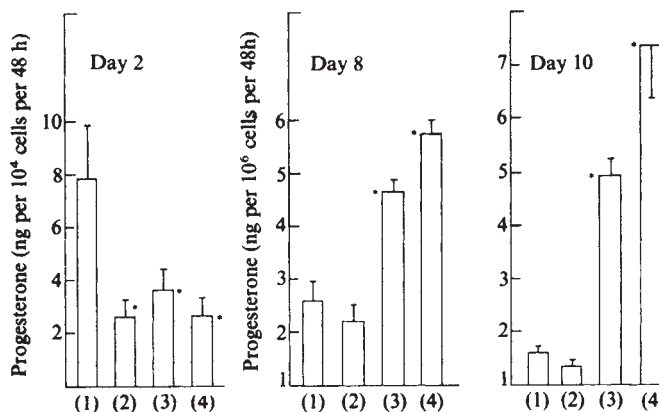


Fig. 2 Acute inhibitory and delayed stimulatory interactions of oestrogen and prolactin on steroidogenesis in cultured porcine granulosa cells. Data are expressed as the mass of progesterone (ng) produced per 10⁴ cells (day 2), or per 10⁶ cells (days 8 and 10). Cultures were treated with ovine prolactin (100 ng ml⁻¹) (column (2)) or 17β-oestradiol (1 μg ml⁻¹) (column (3)) or both (column (4)), for 2, 8 or 10 days before enumeration of cell density by electronic (Coulter) particle counting⁵. Column (1) shows control levels. * $P < 0.01$ versus control.

concentrations increase 20–100-fold during follicular maturation *in vivo*^{19,20}, our observations of oestrogenic effects in cultured porcine granulosa cells suggest that oestrogens may govern the physiological responses of these cells to prolactin *in vivo* as well as *in vitro*.

The mechanism of the potent interaction between oestrogens and prolactin is a challenge of these studies. We previously demonstrated that porcine granulosa cells freshly collected from small immature follicles exhibit high concentrations of specific, high-affinity prolactin binding sites¹⁸, which are maintained during continuous monolayer culture⁵. Oestradiol seems to double prolactin binding by these cultured porcine granulosa cells²¹. However, a simple increase in prolactin receptor concentrations would fail to explain the directional reversal of prolactin effects in oestrogen-treated cells. Conversely, prolactin modification of cellular oestrogen binding must also be considered. For example, in rat luteal tissue *in vivo*, prolactin induces oestradiol receptors²², and is synergistic with oestrogen in enhancing steroidogenesis. However, regulatory changes at other cellular loci of hormone action will also require further examination, as the intracellular mechanisms subserving the actions of prolactin and the steroidogenic effects of oestrogens are poorly understood. Further investigation of the hormonally responsive ovarian cells used in the present studies should allow a more comprehensive understanding of the cellular basis for these critical hormone interactions in the mammalian ovary.

We thank the Hormone Distribution Office, NIAMDD, NIH, for ovine prolactin, the Shamokin Packing Corporation for porcine ovaries, Marlene Thompson for manuscript assistance, and Patricia Klase for technical support in the radioimmunoassays. This work was supported by NIH grant HD10122 from NICHD, and NIH Research Fellowship Service Award 1F32HD05755 to J.D.V.

Received 27 September 1979; accepted 9 January 1980.

- Gibori, G. & Richards, J. S. *Endocrinology* **102**, 767–774 (1978).
- Frantz, A. G. *New Engl. J. Med.* **298**, 201–207 (1978).
- Frantz, A. G., Kleinberg, D. L. & Noel, G. L. *Recent Prog. Horm. Res.* **28**, 527–590 (1972).
- McNatty, K. P., Sawers, R. & McNeilly, A. S. *Nature* **250**, 653–655 (1974).
- Veldhuis, J. D. & Hammond, J. M. *Biol. Reprod.* **20**, Abstr. 118 (1979).
- Veldhuis, J. D. & Hammond, J. M. *Endocrinology* **105**, 1143–1151 (1979).
- Goldenberg, R. L., Vaitukaitis, J. L. & Ross, G. T. *Endocrinology* **90**, 1492–1498 (1972).
- Hilliard, J. *Biol. Reprod.* **8**, 203–214 (1973).
- Bender, E. M., Miller, J. B., Possley, R. M. & Keyes, P. L. *Endocrinology* **103**, 1937–1943 (1978).
- Haney, A. F. & Schomberg, D. W. *Biol. Reprod.* **19**, 242–248 (1978).
- Williams, M. T. & Marsh, J. M. *Endocrinology* **103**, 1611–1618 (1978).
- Akbar, A. M., Stormshak, F. & Lee, D. J. *J. Anim. Sci.* **35**, 1206–1211 (1972).

13. Stouffer, R. L., Nixon, W. E., Gulyas, B. J. & Hodgen, G. D. *Endocrinology* **100**, 506–511 (1977).
14. Williams, M. T., Roth, M. S., Marsh, J. M. & LeMaire, W. J. *J. clin. Endocr. Metab.* **48**, 437–440 (1979).
15. Bernard, J. *J. Reprod. Fert.* **43**, 453–460 (1975).
16. Goldenberg, R. L., Bridson, W. E. & Kohler, P. O. *Biochem. biophys. Res. Commun.* **48**, 101–108 (1972).
17. Thanki, K. H. & Channing, C. P. *Endocrinology* **103**, 74–80 (1978).
18. Rolland, R., Gunsalus, G. L. & Hammond, J. M. *Endocrinology* **98**, 1083–1091 (1976).
19. McNatty, K. P., Hunter, W., McNeilly, A. S. & Sawers, R. S. *J. Endocr.* **64**, 555–571 (1975).
20. Eiler, H. & Nalbandov, A. V. *Endocrinology* **100**, 331–338 (1977).
21. Hammond, J. M. & Krall, E. *Biochem. biophys. Res. Commun.* **91**, 284–288, 1979.
22. Gibori, G., Richards, J. S. & Keyes, P. L. in *Ovarian Follicular and Corpus-Luteum Function*, **112** (eds Channing, C. P., Marsh, J. & Sadler, W. A.) 53–69 (Plenum, New York, 1979).
23. Thorneycroft, I. H. & Stone, S. C. *Contraception* **5**, 129–136 (1972).

Axons from CNS neurones regenerate into PNS grafts

P. M. Richardson, U. M. McGuinness & A. J. Aguayo

The Neurosciences Unit, Montreal General Hospital and McGill University, 1650 Cedar Avenue, Montreal, Canada H3G 1A4

Axons in the peripheral nervous system (PNS) and central nervous system (CNS) form sprouts after injury^{1–3}. Elongation of regenerating axonal sprouts has been observed as the exception within the adult mammalian CNS but is the rule in the PNS of mammals as well as in the CNS of some fish and amphibians⁴. The relative importance of intrinsic neuronal properties and axonal environment in determining the extent of axonal regrowth is unknown⁵. Neuroglial cells, nerve growth factor and target tissues such as smooth muscle are known to influence neuronal responses to injury^{6,7}. Here we have examined the capacity of transected axons originating in the CNS to regrow into nerve grafts containing Schwann cells.

The surgical technique used, a modification of that previously described^{8,9}, was carried out in young adult Sprague–Dawley rats. In the first group of 16 animals, a 5-mm segment of the midthoracic spinal cord was removed and an autologous sciatic nerve graft was inserted sub-pially between the two stumps of the spinal cord. One to four months later, the animals were anaesthetised and perfused. The grafts and adjacent spinal cord tissue were examined by light and electron microscopy. The perineurium surrounding the common peroneal and posterior tibial fascicles of the sciatic nerve was clearly identifiable in cross-sections through the middle of the graft and contained many myelinated or unmyelinated axons ensheathed by Schwann cells. The ultrastructural appearance of each graft was similar to that of a regenerated peripheral nerve. The grafts were in gross continuity with the proximal and distal spinal cord stumps. At the spinal cord–graft interfaces, small cysts were frequently found but were rarely greater than 1 mm in diameter. In electron micrographs, the marginal spinal cord tissue was seen to contain many astrocytic processes and to be surrounded by a basal lamina. Small, irregular dome-like neuroglial structures were observed to protrude towards the graft at the ragged edge of the spinal cord tissue. These protuberances resemble structures seen at the interface of optic nerve grafts and peripheral nerves¹⁰ and also at the normal dorsal root entry zone¹¹. Occasional fibres were found where the axon was surrounded by peripheral myelin (periodicity 15.0 nm) and Schwann cell cytoplasm on one side of a node of Ranvier and central myelin (periodicity 13.6 nm) and astrocytic processes on the other. From these morphological observations we conclude that some axons crossed the border between the graft and spinal cord, although the direction in which they grew could not be determined.

Because of the potential ability of axons in the dorsal and ventral spinal roots to innervate such grafts, all roots crossing the pia-arachnoid at the graft site as well as the corresponding dorsal root ganglia were avulsed in other rats. Grafts in rats with

avulsed dorsal root ganglia contained a mean of 5,850 myelinated axons (s.e. = 510, $n = 6$); those in rats with intact spinal roots and ganglia contained a mean of 10,950 myelinated axons (s.e. = 1,660, $n = 5$). These results suggest that dorsal and ventral roots contributed to the innervation of the grafts but were not the sole source of axons. They do not exclude the possibility that the grafts were entirely innervated by axons originating from dorsal root ganglia and motor neurones¹².

A third group of experiments were carried out to identify the cellular origin of axons in the graft. In 15 rats, a segment of the spinal cord at least 10 mm long was removed and a sciatic nerve graft of equal length was inserted. The roots were left intact. Three to four months later the operative site was re-exposed and horseradish peroxidase (HRP) was injected into the spinal cord immediately rostral or caudal to the graft. After 2 more days, each animal was killed and labelled neurones were sought in a 10-mm segment of the non-injected spinal cord stump, immediately adjacent to the graft (Fig. 1). After injection rostral to the graft, the mean number of labelled neurones seen in the caudal spinal cord stump was 48 (0, 8, 28, 40, 169); after injection caudal to the graft, the mean number in the rostral stump was 6 (0, 0, 1, 6, 23). In five control animals, the graft was crushed with jeweller's forceps immediately before HRP injection and no HRP was seen 2 d later in spinal neurones on the other side of the graft. We conclude that neurones in the experimental groups were not labelled spuriously because of interstitial diffusion through the graft, flow in the cerebrospinal fluid or extravasation into the systemic vasculature. Labelling is thought to have resulted from incorporation of HRP into axon terminals at or near the spinal cord–graft junction and retrograde axonal transport across the graft to perikarya in the adjacent spinal cord. In segments below the graft, labelled

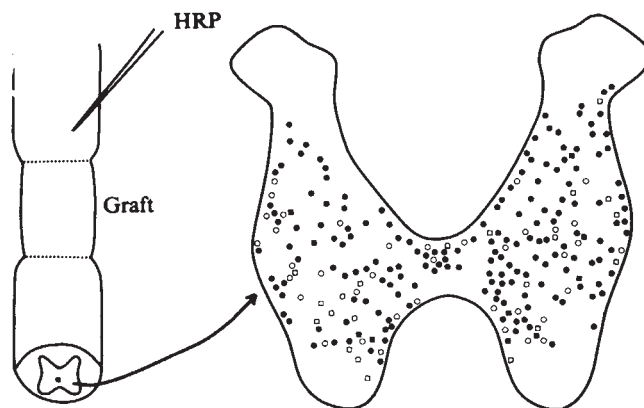


Fig. 1 Left, 2–4 months after the original operation, HRP (20% Sigma VI) was injected through a glass pipette into the spinal cord immediately rostral (or caudal) to the sciatic nerve graft. The injection of small quantities of HRP (0.5 μ l in 30 min) was designed to minimise the possibility of a false positive result due to the extravasation of enzyme and not to label every axon crossing the graft. Vaseline was spread extradurally and subdurally about the graft before injection, dry cotton batting was placed about the pipette during the injection and mineral oil was injected through the pipette as it was withdrawn. Two days later, the animal was perfused with glutaraldehyde. From a 10-mm segment of the spinal cord immediately across the graft from the injection site, specimens were removed and rinsed overnight in sucrose buffer. Approximately 300 sections, each 20 μ m thick, were cut on a freezing microtome, mounted, incubated with tetramethyl benzidine and hydrogen peroxide and counterstained with neutral red²⁴. Right, composite diagram illustrating the position of labelled nerve cell bodies in the spinal grey matter below the graft in four animals after HRP injection above the graft. Each animal is represented by a different symbol. Few neurones are labelled in the tips of the dorsal horns and in the lateral portion of the anterior horns. Because of the small amount of HRP injected and the short length of the spinal cord surveyed, it is assumed that more spinal neurones project into the graft than are shown here.

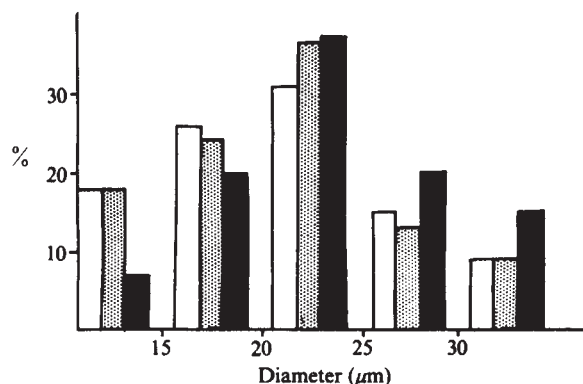


Fig. 2 Histogram showing the diameters as measured under the light microscope, of perikarya in the grey matter of the lower thoracic cord (excluding the dorsal part of the dorsal horn and lateral part of the ventral horn). White columns represent neurones in a normal rat; dotted columns represent unlabelled neurones below the graft; black columns represent labelled neurones below the graft. Labelled neurones have a larger mean diameter than either control group.

neurones were scattered through the central part of the grey matter (laminae 4, 5, 7, 8, 10)¹³ (Fig. 1). Both small and large nerve cells were labelled and the diameter of HRP-containing neurones was slightly greater than that of unlabelled neurones in the same regions of the spinal grey matter (Fig. 2). Some anterior horn cells above the graft and some dorsal root ganglia at the level of the graft and several segments below or above it also contained labelled cells but the ganglia were not surveyed systematically. The evidence demonstrates that some axons originating from neurones within the spinal cord grew a distance of approximately 10 mm into a PNS graft. We do not know whether such axons then re-entered the spinal cord and formed synapses. No functional improvement attributable to axonal regeneration was observed in any animal.

Mammalian central neurones with recognised capacity for axonal regrowth over a distance can be placed in two groups. In the first group, which includes somatic motor neurones and autonomic preganglionic neurones, the axons project outside the CNS. Intrinsic CNS neurones with regenerative ability have been found in the hypothalamus¹⁴ and brain stem^{15,16}. The latter neurones are cholinergic or monoaminergic and are thought to have thinly myelinated or unmyelinated axons¹⁷. The findings reported here indicate that there are other neurones in the rat spinal cord also capable of axonal elongation after injury. Some of the HRP-labelled perikarya probably represent autonomic preganglionic neurones about the central canal¹⁸ or in the intermediolateral columns, but most of them were found in areas of spinal grey matter normally occupied by neither autonomic nor somatic motor neurones. None is likely to have been monoaminergic because cell bodies containing catecholamines or serotonin have not been discovered in the spinal grey matter¹⁹. In short, the labelled cells do not belong to a class of neurones in which axonal regrowth has been described previously. It is unknown whether or not spinal axons with regenerative potentiality have any specific anatomical or biochemical property.

Spinal axons such as those which grew into a peripheral nerve graft in these experiments have repeatedly demonstrated only abortive sprouting after simple transection^{1,20}. This evidence that the regenerative response of similar axons differs in CNS and PNS neuroglia, together with other experimental observations^{10,21,22}, supports the hypothesis that Schwann cells are more conducive to axonal regeneration than central neuroglial cells²³. The regenerative potentiality of CNS neurones may be expressed only when their neuroglial environment is changed.

This work was supported by the MRC of Canada, the Multiple Sclerosis Society of Canada and the Montreal General Hospital Research Institute.

Received 14 September 1979; accepted 21 January 1980.

1. Ramon Y Cajal, S. *Degeneration and Regeneration of the Nervous System* (ed. May, R. M.) (Oxford University Press, London, 1928).
2. Liu, C. N. & Chambers, W. W. *Archs Neurol. Psychiat.* **79**, 46-61 (1958).
3. Raisman, G. *Ann. Neurol.* **3**, 101-106 (1978).
4. Singer, M., Nordlander, R. H. & Egar, M. J. *comp. Neurol.* **185**, 1-22 (1979).
5. Grafstein, B. & McQuarrie, I. G. in *Neuronal Plasticity* (ed. Cotman, C. W.) (Raven, New York, 1978).
6. Stenevi, U., Bjerre, B., Bjorklund A. & Mobley, W. *Brain Res.* **69**, 217-234 (1974).
7. Varon, S. S. & Bunge, R. P. *A. Rev. Neurosci.* **1**, 327-361 (1978).
8. Sugar, O. & Gerard, R. W. *J. Neurophysiol.* **3**, 1-19 (1940).
9. Kao, C. C., Chang, L. W. & Bloodworth, J. M. B. *Expl Neurol.* **54**, 591-615 (1977).
10. Aguayo, A. J. *et al. Neurosci. Lett.* **9**, 97-104 (1978).
11. Berthold, C. H. & Carlstedt, T. *Acta physiol. scand. suppl.* **446**, 23-42 (1977).
12. Bratten, B. & Hudson, A. *Can. J. Neurol. Sci.* **6**, 394-395 (1979).
13. Rexed, B. *J. comp. Neurol.* **100**, 297-379 (1954).
14. Adams, J. H., Daniel, P. M. & Prichard, M. M. L. *J. comp. Neurol.* **135**, 121-144 (1969).
15. Katzman, R., Broida, R. & Raine, C. S. *Brain Res.* **138**, 423-443 (1977).
16. Svendgaard, N. A., Bjorklund, A. & Stenevi, U. *Adv. Anat. Embryol. cell. Biol.* **51**, pt 4, 7-77 (1975).
17. Svendgaard, N. A., Bjorklund, A. & Stenevi, U. *Brain Res.* **102**, 1-22 (1976).
18. Hancock, M. B. & Peveto, C. A. *J. comp. Neurol.* **183**, 65-72 (1979).
19. Carlsson, A., Falck, B. & Hillarp, N. A. *Acta physiol. scand.* **60**, 112-119 (1964).
20. Lampert, P. & Cressman, M. *Lab. Invest.* **13**, 825-839 (1964).
21. Weinberg, E. L. & Spencer, P. S. *Brain Res.* **162**, 273-279 (1979).
22. Stensaas, L. J., Burgess, P. R. & Horsch, K. W. *Soc. Neurosci. Abstr.* **5**, 684 (1979).
23. Aguayo, A. J., Bray, G. M., Perkins, C. S. & Duncan, I. D. in *Soc. Neurosci. Symp. IV* (ed. J. A. Ferrendelli) 361-383 (1979).
24. Mesulam, M. M. *J. Histochem. Cytochem.* **24**, 1273-1280 (1978).

Response to stress of mesocortico-frontal dopaminergic neurones in rats after long-term isolation

G. Blanc, D. Hervé, H. Simon*, A. Lisoprawski, J. Glowinski & J. P. Tassin

Groupe NB, INSERM U.114, Collège de France, 11, place Marcelin Berthelot, 75231 Paris Cedex 05, France

Recent studies suggest that the mesocortico-frontal dopaminergic neurones which originate in the ventral tegmental area (VTA) have an inhibitory role in locomotor activity^{1,2}. They are also markedly activated under stress³⁻⁵. This effect was shown in rats and mice subjected to electric foot-shocks by measuring either the rate of decline of dopamine (DA) after α -methylparatyrosine treatment³ or the changes in dihydroxyphenylacetic acid (dopac) levels and the dopac/DA ratio^{4,6}. In rats, stress-induced activation of the dopaminergic neurones was prevented by benzodiazepines^{4,5}, and studies in BALB/c mice introduced for 2 min into an open field further established the role of dopaminergic neurones in emotional responses⁷. These observations led us to examine the effects of long-term isolation on the activity of the mesocortico-frontal dopaminergic neurones in rats, some of which were subjected to a stressful situation. Indeed, several groups have reported that long-term isolation in rodents induced behavioural disturbances such as increased motor activity^{8,9} and aggression^{9,10} and hyper-reactivity to a new environment or stressful stimuli^{10,11}. As measured by the changes in dopac levels or the dopac/DA ratio, we report here that the activity of the mesocortico-frontal dopaminergic neurones was reduced after isolation. This was not the case for the dopaminergic neurones projecting to the nucleus accumbens or the striatum, the rate of DA utilisation in these structures was even enhanced in isolated rats in which the activity of the mesocortico-frontal dopaminergic neurones was markedly reduced. Finally, we will show that a 3-min electric foot-shock session is more effective in enhancing dopac levels or the dopac/DA ratio in the frontal cortex of isolated than grouped rats.

*Laboratoire de Neurobiologie du Comportement, Université de Bordeaux II, Rue Le's Saignat, 33076 Bordeaux Cedex, France.

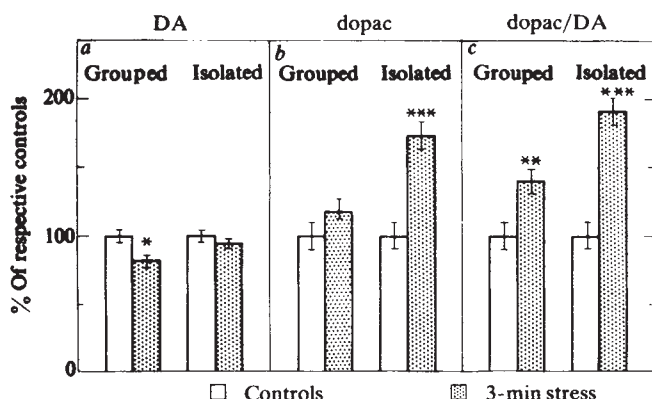


Fig. 1 Effect of stress on the rate of dopamine (DA) utilisation in the frontal cortex of grouped or isolated rats. Rats were isolated or kept in groups for 8 or 12 weeks and then individually exposed to a 3-min electric foot-shock stress. Other isolated or grouped rats were used as respective controls. In all cases, DA (a) and dopac (b) levels were estimated in the frontal cortex and the dopac/DA ratio (c) was calculated for each sample. Results are the mean of data from two independent experiments in which similar changes were observed. They are thus the mean $\pm$ s.e.m. of data obtained with 20 samples (10 animals in each case) and are expressed in per cent of respective control values. The mean values of DA, dopac and dopac/DA ratios in grouped and isolated rats, respectively, are 1,150, 325, 0.29 and 1,250, 300, 0.26, DA and dopac levels being expressed in pg per mg of protein. $\square$, Controls; $\square$ 3-min stress. * $P < 0.05$; ** $P < 0.02$; *** $P < 0.001$ when compared with respective control values.

Male Charles River rats (180–200 g) were kept in a controlled environment (24 °C, 60% humidity, alternate cycles of 12 h light (0700–1900 h) and darkness) and received food and water *ad libitum*. Some of them were isolated in opaque cages (28 $\times$ 18.5 $\times$ 12 cm) for at least 8 weeks, others remained grouped (46 $\times$ 26 $\times$ 20 cm), (five animals per cage). The isolated rats could hear and smell their congeners, but were unable to see, touch or otherwise interact with them. Isolated and grouped rats were killed the same day between 1100 and 1200 h. Their brains were immediately frozen and cut in 500- μ m coronal slices from which tissue samples were punched out in both frontal cortices (11,300 and 10,800 μ m slices)⁴. In some experiments, samples were also taken bilaterally from the nucleus accumbens (9,800 and 9,300 μ m) and the medial part of the striatum⁴ (8,800 μ m) which is mainly innervated by nigral dopaminergic cells. After separation on alumina microcolumns, DA and dopac levels were estimated with a radioenzymatic method and the dopac/DA ratio was determined for each sample^{4,12}.

From the results of five independent experiments carried out during a 10-month period, the mean dopac/DA ratio was found to have decreased in frontal cortices of isolated rats (0.24 ± 0.010 , $n = 50$ samples, -30% , $P < 0.001$) when compared with that of grouped animals (0.34 ± 0.014 , $n = 50$). This effect, which was mainly due to a decrease in dopac levels (-19% , $P < 0.01$), suggests that isolation reduced the activity of the mesocortico-frontal dopaminergic neurones. However, the decreased dopac/DA ratio observed in isolated rats varied between experiments (-4% , -19% , -32% , -35% , -43%), the effect being significant in three cases. This seemed to be related to variations in the mean values of the dopac/DA ratio in the grouped rats (0.27–0.41); the ratio was remarkably constant from one experiment to another in isolated animals (0.23–0.26). Seasonal rhythms¹³ and/or seasonal changes in emotional states could be responsible for the variations observed in grouped rats. Prolonged isolation may attenuate these fluctuations.

In two of the five experiments, half of the rats isolated for 8 weeks were grouped for 4 additional weeks. In this case, their cortical dopac/DA ratio was significantly higher than that estimated in rats isolated for 12 weeks, but not different from that of rats continuously kept in groups (Table 1). Therefore, the reduced activity of the mesocortico-frontal DA neurones induced by isolation is a reversible process.

In two experiments we examined simultaneously the reactivity of the mesocortico-frontal DA neurones in grouped and isolated rats exposed for 3 min in individual cages to electric foot-shocks⁴. Confirming previous results, this stress reduced the levels of DA and increased the dopac/DA ratio in the frontal cortices of grouped rats⁴. The effect was more pronounced in isolated animals, the increase in the dopac/DA ratio being twice that observed in grouped rats and mainly due to an increase in dopac level ($+73\%$) (Fig. 1). Because the absolute levels of dopac and the dopac/DA ratio were higher in isolated (519 ± 45 pg per mg protein, 0.50 ± 0.04 , $P < 0.05$) than in grouped (383 ± 36 pg per mg protein, 0.40 ± 0.03) rats, the enhanced reactivity of the DA neurones seen under stress in isolated animals cannot be attributed solely to their initial lower state of activity. The biochemical changes observed in isolated rats, which were identical to those previously seen in grouped animals exposed to a longer stress session (20 min)⁴ suggest that the intense activation of the DA neurones enhanced not only the rate of DA utilisation but also its synthesis. Behavioural differences were also observed: the grouped rats jumped during electric shocks and attempted to escape, whereas the isolated rats exhibited stereotypic and freezing behaviours during and between shocks, respectively. The grouped rats were highly reactive at the end of the 3-min stress whereas the isolated animals seemed exhausted.

Finally, in three of the five experiments we also compared the rate of DA utilisation in the nucleus accumbens and the striatum of isolated and grouped rats. The mean results show a slight but not significant enhanced utilisation of DA in these structures in isolated rats. This finding is in contrast with the effect observed in the frontal cortex. Moreover, in one experiment in which a pronounced reduction of the dopac/DA ratio was seen (-43%), significant opposite effects were obtained in the nucleus accumbens ($+36\%$) and striatum ($+33\%$) (Fig. 2).

Changes in monoamine metabolism in the whole brain of isolated rats or mice have already been reported. Despite some discrepancies, most authors found that the turnover rate of serotonin, noradrenaline and DA were reduced in isolated rats^{8–11}. In one study, DA turnover was decreased in the amygdala but was slightly increased in the olfactory tubercles and remained unchanged in the striatum and the nucleus accumbens¹⁴.

Our results extend previous findings and demonstrate that long-term deprivation of social interaction in rats reduced and

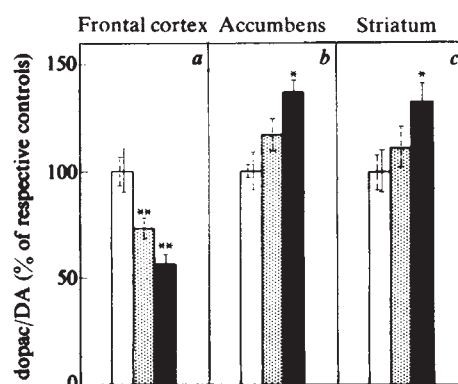


Fig. 2 Effects of isolation on the dopac/DA ratio in the frontal cortex, nucleus accumbens and striatum. Rats were isolated or kept in groups for 8 or 12 weeks and the dopac/DA ratio was estimated bilaterally in areas of the frontal cortex (a), nucleus accumbens (b) and striatum (c). For each structure, the results are the mean $\pm$ s.e.m. of 30 samples (15 rats) and 10 samples (5 rats), respectively. They are expressed in per cent of corresponding values: 0.33, 0.15, 0.09 (30 samples) and 0.41, 0.18, 0.12 (10 samples) for the frontal cortex, nucleus accumbens and striatum, respectively. $\square$, Grouped; $\square$, isolated, mean of three independent experiments; $\blacksquare$, isolated, values of one of three experiments in which the maximal decrease in the cortical dopac/DA ratio was observed.

* $P < 0.02$, ** $P < 0.001$ when compared with respective control values.

Table 1 Reversibility of the effects of long-term isolation on the dopac/DA ratio and dopac levels in the rat frontal cortex

	dopac/DA	dopac (pg per mg protein)
Grouped (12 weeks)	0.40 ± 0.024	430 ± 35
Isolated (12 weeks)	0.24 ± 0.010	309 ± 19
	(-40%)*	(-28%)+
Isolated (8 weeks) then grouped (4 weeks)	0.31 ± 0.017	401 ± 24
	(+29%)+	(+30%)+

Rats were isolated or kept in groups (five per cage) for 12 weeks. Other animals were first isolated for 8 weeks and then grouped for 4 weeks. DA and dopac levels were estimated bilaterally in the frontal cortex and the dopac/DA ratio was calculated for each sample. Results are the mean of data obtained in two independent experiments in which similar changes were observed. Thus, values of the dopac/DA levels are the mean ± s.e.m. of data obtained with 20 samples (10 animals) in each case.

* $P < 0.001$; + $P < 0.01$ when compared with values obtained with grouped animals (Student's *t*-test). ‡ $P < 0.01$ when compared with values obtained with rats isolated for 12 weeks.

maintained at a low level the activity of their mesocortico-frontal dopaminergic neurones. Indeed, the cortical dopac/DA ratio was constant during a 10-month period and was lower than that observed in grouped animals. Furthermore, isolation enhanced the reactivity of the mesocortico-frontal dopaminergic neurones to stress, as a 3-min session of electric foot-shocks was more effective in enhancing the cortical dopac levels and dopac/DA ratio in isolated than in grouped animals. This suggests that under stress the transmission of messages responsible for the activation of the mesocortico-frontal dopaminergic neurones is facilitated by previous isolation.

In recent studies, we showed that mesocortico-frontal dopaminergic neurone activity was controlled by interneuronal regulations which differed from those of other VTA dopaminergic cells^{15,16}. The present results confirm these findings because, in contrast to that observed for the mesocortico-frontal dopaminergic neurones, the activity of the dopaminergic neurones innervating the nucleus accumbens was not reduced in isolated rats. In fact, the opposite occurred in both the nucleus accumbens and the striatum, when the cortical dopac/DA ratio was reduced to a large extent in isolated rats. These results agree with Carter and Pycock's recent observations suggesting that the mesocortico-frontal dopaminergic neurones exert an inhibitory role on the activity of dopaminergic neurones innervating subcortical structures². Indeed, these authors have shown that the specific destruction of fronto-cortical dopaminergic terminals in the rat not only enhanced the turnover of DA in the nucleus accumbens and striatum but also increased the locomotor activity and stereotypic effect of amphetamine².

We already know that the mesocortico-frontal dopaminergic neurones are involved in the control of locomotor activity^{7,2}, emotional behaviour^{4,17} and cognitive processes^{18,19}. Therefore, the prolonged reduced activity of the mesocortico-frontal dopaminergic neurones in isolated rats as well as their higher reactivity under stress could be responsible for the enhanced locomotor activity and hyperactivity seen in these animals.

We thank Mr G. Godeheu for his help in taking care of the animals. This study was supported by grants from INSERM (CL 79.5.362.6), DRET (79.077) and Rhône-Poulenc SA.

Received 17 December 1979; accepted 24 January 1980.

1. Tassin, J. P. *et al.* *Brain Res.* **141**, 267–281 (1978).
2. Carter, C. J. & Pycock, D. J. *Br. J. Pharmacol.* **64**, 430P (1978).
3. Thierry, A. M., Tassin, J. P., Blanc, G. & Glowinski, J. *Nature* **263**, 242–244 (1976).
4. Lavielle, S. *et al.* *Brain Res.* **168**, 585–594 (1979).
5. Fadda, F. *et al.* *Life Sci.* **27**, 2219–2224 (1978).
6. Hervé, D. *et al.* *Life Sci.* **25**, 1659–1664 (1979).
7. Tassin, J. P., Hervé, D., Blanc, G. & Glowinski, J. *Neurosci. Lett.* (in the press).
8. Weinstock, M., Speiser, Z. & Ashkenazi, R. *Psychopharmacology* **56**, 205–209 (1978).
9. Valzelli, L. *Psychopharmacologia* **31**, 305–320 (1973).
10. Welch, A. S. & Welch, B. L. 91–142 (Plenum, New York, 1971).
11. Brain, P. & Benton, D. *Life Sci.* **24**, 99–116 (1979).

12. Gauchy, C., Tassin, J. P., Glowinski, J. & Chéramy, A. *J. Neurochem.* **26**, 99–104 (1976).
13. Sheving, L. E., Harrison, W. H., Gordon, N. P. & Pauly, J. E. *Am. J. Physiol.* **214**, 166–173 (1968).
14. Thoa, N., Tizabi, Y. & Jacobowitz, D. *Brain Res.* **131**, 259–269 (1977).
15. Lisoprawski, A., Hervé, D., Blanc, G., Glowinski, J. & Tassin, J. P. *Brain Res.* **183**, 229–234 (1980).
16. Hervé, D. *et al.* *Neurosci. Lett.* **15**, 127–133 (1979).
17. LeMoal, M. *et al.* *Adv. biochem. Psychopharmacol.* **16**, 237–245 (1977).
18. Brozoski, T. S., Brown, R. M., Rosvold, H. E. & Goldman, P. S. *Science* **205**, 929 (1979).
19. Simon, H., Scatton, B. & Le Moal, M. *Neurosci. Lett.* **15**, 319–324 (1979).

Cycloheximide-sensitive synthesis of substance P by isolated dorsal root ganglia

A. Harmar*, J. G. Schofield & P. Keen

Departments of Pharmacology and Biochemistry, Medical School, University Walk, Bristol BS8 1TD, UK

Substance P (Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH₂) may be used as a neurotransmitter by certain primary afferent neurones^{1,2}, particularly those carrying pain impulses^{3–7}. Substance P-like immunoreactivity has been localised to the cell bodies of one population of dorsal root ganglion neurones by immunocytochemistry⁸. It is contained in vesicles^{9,10} in the central terminals of these neurones⁸, and has also been demonstrated in the peripheral terminals^{11,12}. As axons and terminals have very little capacity for peptide biosynthesis¹³, it is possible that substance P is synthesised and packaged in the perikaryon and transported to the terminals by an axoplasmic transport process. Consistent with this is the finding that substance P accumulates proximal to a ligature placed on the dorsal root¹⁴. There has, however, been no direct demonstration of the biosynthesis of substance P in the nervous system. We report here that rat dorsal root ganglia incorporate ³⁵S-methionine into substance P, characterised as authentic by immunoprecipitation followed by HPLC. There is a delay of 1–2 h between addition of label and its incorporation into substance P. Synthesis is blocked by cycloheximide suggesting that, in dorsal root ganglia, substance P is synthesised by a conventional ribosomal process. Synthesis of substance P is reduced by some 90% in ganglia from rats treated neonatally with capsaicin, a drug which is thought to destroy a population of primary afferent neurones¹⁵.

The anti-substance P serum was raised in rabbits using synthetic substance P (UCB Bioproducts) conjugated to succinylated thyroglobulin by the carbodiimide method. As expected with this coupling reagent, which reacts with the amino groups on residues 1 and 3 of substance P, the antiserum was directed to the C-terminal region of the peptide. The affinity of a number of substance P analogues for the antiserum was tested by radioimmunoassay using ¹²⁵I-Tyr⁸ substance P as tracer. The affinities of fragment 4–11, the substance P analogue H-Lys-Phe-Tyr-Gly-Leu-Met-NH₂ (CGP 15 898) and of substance P-free acid, which differs from substance P only in that it lacks the C-terminal carboxamide group, were respectively 2-, 10- and 4,000-fold less than that of the parent peptide.

Rat lumbar dorsal root ganglia were incubated with ³⁵S-methionine, homogenised, centrifuged and the radiolabelled peptides in the supernatant immunoprecipitated with anti-substance P serum. The immunoprecipitates were washed, redissolved by boiling in HPLC buffer and the radiolabelled peptides separated by reverse-phase HPLC (see legend to Fig. 1 for experimental details). In initial experiments designed to establish the specificity of the immunoprecipitation, washed immunoprecipitates were incubated overnight with a large excess (50 µg) of substance P in buffer A (see legend to Fig. 1). After centrifugation the supernatant was diluted with 1 ml buffer B and the labelled peptides separated by HPLC in the usual way. Elution by displacement with substance P and by

* To whom correspondence should be addressed.

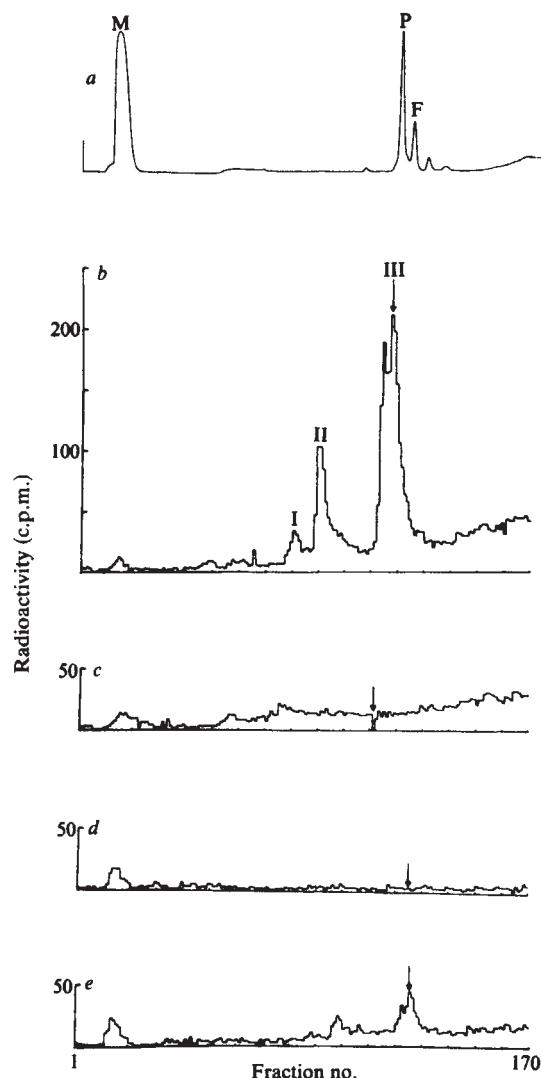


Fig. 1 HPLC of immunoprecipitates from rat dorsal root ganglia incubated with ^{35}S -methionine for 9 h. *a*, Standards detected by UV absorption. M, methionine; P, substance P; F, substance P fragment 4–11. *b–e*, distribution of radioactivity. Arrow denotes position of carrier substance P: *b*, precipitated with immune serum; *c*, same but precipitated with pre-immune serum; *d*, as *b* but 0.1 mM cycloheximide included in incubation medium; *e*, dorsal root ganglia from rat which had been treated neonatally with capsaicin. Rats were perfused with ice-cold Krebs–bicarbonate. Dorsal root ganglia L4–6 were dissected out, placed in 0.5 ml of incubation medium containing (mM) NaCl, 118; KCl, 4.5; CaCl_2 , 2.5; MgCl_2 , 1.15; HEPES (pH 7.5), 15; glucose 28 and 100 μCi ^{35}S -labelled L-methionine (Radiochemical Centre, 710 Ci mmol^{-1}), incubated with shaking at 37 °C under O_2 for 9 h, washed, held in 200 μl 2M acetic acid in a boiling waterbath for 10 min and homogenised. In experiments with cycloheximide ganglia were pre-incubated for 1 h with 0.1 mM cycloheximide, ^{35}S -methionine was then added and incubation was allowed to proceed for a further 9 h in the continued presence of cycloheximide. The homogenate was centrifuged and the supernatant lyophilised. The lyophilised samples were taken up in 250 μl buffer A (50 mM barbitol buffer, pH 8.6 containing 1 mM EDTA, 0.5 mg ml^{-1} L-methionine and 0.1% Triton X-100) and centrifuged. Aliquots (110 μl) of the supernatant were incubated for 16 h on ice with 20 μl buffer A (to which had been added 20 mg ml^{-1} bovine serum albumin and 50 $\mu\text{g ml}^{-1}$ poly-lysine) and 10 μl anti-substance P serum or pre-immune serum. 14 μl goat anti-rabbit IgG serum (Miles) and 26 μl buffer A were then added and immunoprecipitates were allowed to form for a further 4 h at 4 °C. All sera were pretreated with phenylmethyl sulphonyl fluoride and iodoacetamide to inhibit protease activity²⁷. Immunoprecipitates were washed three times with 1 ml buffer A then redissolved in 1 ml buffer B (0.1 M NaH_2PO_4 , 0.1 M H_3PO_4 , pH 2.1) containing 50 μg substance P, held in a boiling waterbath for 10 min and injected on to a 250 $\times$ 4.6 mm column of Partisil 10 ODS using a 2 ml sample loop. After a loading phase of 3 min in buffer B the peptides were eluted with a gradient of acetonitrile in buffer B essentially as described by O'Hare and Nice²⁰ and terminating when the acetonitrile concentration reached 60% (v/v). Carrier substance P was detected by UV absorption at 225 nm and 10-drop fractions were collected for liquid scintillation counting in a Triton X-100 toluene-based scintillation fluid at 80% efficiency.

direct dissolution in HPLC buffer gave similar results, so, as the latter procedure was simpler, it was used routinely.

The ^{35}S -labelled material precipitated by anti-substance P serum separated as three well-defined peaks, one of which (peak III) coincided with authentic carrier substance P (Fig. 1*b*). None of these peaks corresponded with substance P sulfoxide which elutes at a point intermediate between peaks II and III. There was a lag phase before ^{35}S -methionine became incorporated into the substance P peak; no incorporation was detectable after 1 h but it became apparent at 2 h and continued to rise at 3 h and 8 h (Fig. 2). Incorporation of label into peak II followed a similar time course, whereas the radioactivity incorporated into peak I had reached a maximum at 2 h and thereafter remained essentially constant. There was no radioactivity in any of the three peaks when the anti-substance P serum in the immunoprecipitation step was replaced by serum taken from the same animal before immunisation (Fig. 1*c*). Incorporation of radiolabel into substance P was completely blocked by inclusion of 0.1 mM cycloheximide in the incubation medium (Fig. 1*d*).

In ganglia from 10-week-old rats which had been given 0.7 mg capsaicin on the second day of life¹⁵, incorporation into substance P (peak III) and into peaks I and II was respectively 12, 8 and 14% of that in normal animals (Fig. 1*e*). Neonatally-administered capsaicin causes degeneration of a population of small-diameter chemosensitive neurones^{15,16} which are thought to use substance P as a neurotransmitter³. Animals treated in this way have severely reduced levels of substance P in the dorsal horn of the spinal cord¹⁷ and in their dorsal root ganglia (unpublished observation). Morphometric studies of the ganglia and dorsal roots of these animals show that the population of small dark cells and of non-myelinated fibres is severely depleted while the population of myelinated fibres is slightly depleted¹⁸.

The substance P peak and also peak II were re-chromatographed on a 58 $\times$ 1 cm column of Sephadex G-50 in phenol:acetic acid:water 1:1:1, containing 0.5% mercaptoethanol¹⁹. The substance P peak eluted as a single peak on Sephadex G-50 which coincided with ^{125}I -Tyr⁸ substance P, confirming its identity with authentic substance P. Peak II eluted later than ^{125}I -Tyr⁸ substance P on Sephadex G-50, suggesting that it is smaller than substance P. To test the possibility that peak II could have been formed artefactually from ^{35}S -substance P during extraction and chromatography, we added ^{125}I -Tyr⁸ substance P to a ganglion homogenate and carried it through the complete procedure. The ^{125}I -Tyr⁸ substance P treated in this way gave no extra peaks and so it seems likely that both peaks I and II represent peptides synthesised in the tissue. If peak II is a fragment of substance P it must be a C-terminal fragment as it contains ^{35}S -methionine. The retention time of peptides on reverse-phase supports has been shown to be broadly correlated with their hydrophobic amino acid content²⁰. Thus fragments of substance P containing phenylalanine residues at both positions 7 and 8 may be expected to have similar or longer retention times than substance P. For example, fragment 4–11 is retained longer than substance P (Fig. 1*a*). As peak II elutes before substance P it could be a fragment in which one or both the phenylalanine residues has been lost. We are currently attempting to elucidate its structure.

The complete inhibition of substance P biosynthesis by cycloheximide (Fig. 1*d*) suggests that the peptide is synthesised by a conventional ribosomal mechanism. Substance P is released from nerve terminals^{7,21,22} and so it is to be expected that, in common with other exported peptides^{23,24}, substance P is synthesised on membrane-bound ribosomes. Consistent with this is the observation that the perikarya of the small dark neurones, in which substance P-like immunoreactivity is localised⁸ and which are destroyed by capsaicin¹⁵, are replete with rough endoplasmic reticulum²⁵. As it is thought that all peptides synthesised on membrane-bound ribosomes must be at least 40–60 residues in length and bear a hydrophobic N-terminal signal sequence²⁶ one would expect the undecapeptide substance P to be formed from a larger precursor. The latent period of 2 h between the addition of ^{35}S -methionine and the

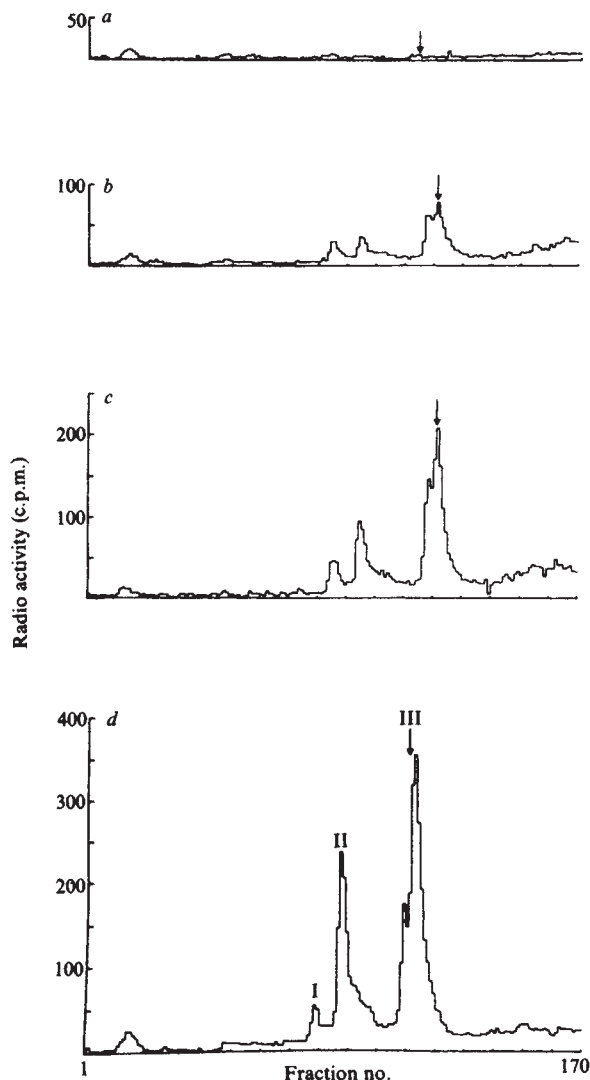


Fig. 2 HPLC of immunoprecipitates from rat dorsal root ganglia incubated with ^{35}S -methionine for 1, 2, 3 and 8 h. Method as in legend to Fig. 1 except that the entire supernatant from each homogenate was reacted with anti-substance P serum. Arrows denote position of carrier substance P.

appearance of the label in substance P (Fig. 2) may represent the time necessary for the processing of such a precursor. Note that the radioactivity incorporated into peak I had reached a maximum at 2 h and thereafter remained constant while incorporation into substance P continued to rise (Fig. 2). This is the behaviour expected of a precursor. Unfortunately the radioactivity incorporated into peak I in these experiments was insufficient to allow us to determine its size by subsequent chromatography on Sephadex G-50.

The present demonstration that isolated rat dorsal root ganglia are able to incorporate ^{35}S -methionine into substance P provides a system in which the synthesis and packaging of neurotransmitter substance P may be examined in detail. Using this system we have obtained evidence that substance P synthesis in dorsal root ganglia occurs by a ribosomal mechanism and that capsaicin may be a useful tool for the selective destruction of neurones responsible for this biosynthesis. We are continuing to investigate the possibility that substance P may be synthesised from larger molecular weight precursors.

We thank Ms S. Hyde for technical assistance, Dr E. Atherton for a gift of synthetic substance P and Dr H. Bittiger of Ciba-Geigy for a gift of CGP 15 898. This work is supported by a project grant from the MRC.

Received 19 November 1979; accepted 31 January 1980.

1. Lembeck, F. *Arch. exp. Path. Pharmacol.* **219**, 197–213 (1953).
2. Otsuka, M. & Konishi, S. *Cold Spring Harb. Symp. quant. Biol.* **40**, 135–143 (1976).
3. Jessell, T. M., Iversen, L. L. & Cuello, A. C. *Brain Res.* **152**, 183–188 (1978).
4. Henry, J. L. *Brain Res.* **114**, 439–451 (1976).
5. Randić, M. & Miletić, V. *Brain Res.* **128**, 164–169 (1977).
6. Hökfelt, T., Ljungdahl, A., Terenius, L., Elde, R. & Nilsson, G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3081–3085 (1977).
7. Jessell, T. M. & Iversen, L. L. *Nature* **268**, 549–551 (1977).
8. Hökfelt, T. *et al. Neuroscience* **1**, 131–136 (1976).
9. Cuello, A. C., Jessell, T. M., Kanazawa, I. & Iversen, L. L. *J. Neurochem.* **29**, 747–751 (1977).
10. Barber, R. P., Slemmon, J. R., Salvatena, P. M., Roberts, E. & Leeman, S. E. *J. comp. Neurol.* **184**, 331–352 (1979).
11. Hökfelt, T., Kellerth, J.-O., Nilsson, G. & Pernow, B. *Brain Res.* **100**, 235–252 (1975).
12. Olgart, L., Hökfelt, T., Nilsson, G. & Pernow, B. *Pain* **4**, 153–159 (1977).
13. Koenig, E. in *Biochem. Simple Neuronal Models* (eds Costa, E. & Giacobini, E.) 303–315 (Raven, New York, 1970).
14. Takahashi, T. & Otsuka, M. *Brain Res.* **87**, 1–11 (1975).
15. Jancsó, G., Kiraly, E. & Jancsó-Gabor, A. *Nature* **270**, 741–743 (1977).
16. Jancsó, G. & Knyihár, E. *Neurobiology* **5**, 42–43 (1975).
17. Gamse, R., Holzer, P. & Lembeck, F. *Br. J. Pharmacol.* **68**, 207–213 (1980).
18. Lawson, S. N. & Nickels, S. M. *J. Physiol.* (in the press).
19. Carnegie, P. R. *Nature* **206**, 1128–1130 (1965).
20. O'Hare, M. J. & Nice, E. *J. Chromatogr.* **171**, 209–226 (1979).
21. Otsuka, M. & Konishi, S. *Nature* **264**, 83–84 (1976).
22. Olgart, L., Gazelius, B., Brodin, E. & Nilsson, G. *Acta physiol. scand.* **101**, 510–512 (1977).
23. Palade, G. *Science* **189**, 347–358 (1975).
24. Shore, G. C. & Tata, J. R. *Biochim. biophys. Acta* **472**, 197–236 (1977).
25. Lieberman, A. R. in *The Peripheral Nerve* (ed. Landon, D. N.) 188–278 (Chapman & Hall, London, 1976).
26. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
27. Mains, R. E. & Eipper, B. A. *J. biol. Chem.* **251**, 4115–4120 (1976).

Enkephalin-, VIP- and substance P-like immunoreactivity in the carotid body

J. Wharton, J. M. Polak*, A. G. E. Pearse, G. P. McGregor, M. G. Bryant, S. R. Bloom, P. C. Emson†, G. E. Bisgard‡ & J. A. Will‡

Departments of Histochemistry and Medicine, Royal Postgraduate Medical School, London W12 0HS, UK

† MRC Neurochemical Pharmacology Unit, Department of Pharmacology, Medical School, Cambridge CB2 2QD, UK

‡ Department of Veterinary Science, University of Wisconsin, Madison, Wisconsin 53706

The carotid body type I cell contains amines^{1–3} and has features, both morphological and cytochemical, which indicate that it may also produce a peptide^{4–6}. Many regulatory peptides are now known to be present in both central and peripheral tissues⁷. In the periphery these neuropeptides occur in both classical endocrine (APUD) cells and the neurones of the autonomic nervous system⁸. We have now investigated the possible presence of neuropeptides in the cat carotid body using both immunocytochemistry and radioimmunoassay. Met- and Leu-enkephalin-like material occurred in considerable quantities in carotid body extracts and enkephalin-like immunoreactivity was localised in type I cells. Both vasoactive intestinal polypeptide (VIP)- and substance P-like immunoreactivity was also present but was localised in nerve fibres distributed throughout the organ. These active neuropeptides are widely distributed in mammalian tissues, forming a diffuse regulatory system which now seems to include the carotid body.

Antisera to Met-, Leu-enkephalin, VIP and substance P were raised in rabbits and characterised for both radioimmunoassay (Table 1) and immunocytochemistry (Table 2). The carotid bodies and bifurcation of the common carotid arteries were removed from adult cats ($n = 25$) under pentobarbitone anaesthesia.

For immunocytochemistry, two methods of fixation were used. (1) The whole bifurcation, including the carotid body, was frozen in melting freon (Arcton), freeze dried, fixed in benzoquinone vapour for 3 h at 60 °C and embedded in paraffin wax⁹.

* To whom reprint requests should be addressed.

Table 1 Characterisation of antisera for enkephalin, VIP and substance P radioimmunoassays

Antiserum to	Carrier and coupler	Label	Sensitivity	Titre	Characterisation
Met-enkephalin	Bovine serum albumin (BSA) by glutaraldehyde	³ H-Met enkaphalin	20–0.025 pmol <0.2% cross reactivity with Leu-enkephalin	1:1,500	C-terminal
VIP	BSA by carbodiimide	¹²⁵ I Chloramine-T method Pure porcine VIP	1 fmol per assay tube	1:320,000	C-terminal
Substance P	BSA or α -globulin by glutaraldehyde	¹²⁵ I Chloramine-T method Tyr ⁸ substance P	0.3 fmol per assay tube	1:8,000	C-terminal

(2) Tissue for liquid fixation was immersed in a 0.4% solution of benzoquinone in 0.01 M phosphate-buffered normal saline (pH 7.1–7.4) for up to 30 min at 4 °C (ref. 10). It was also rinsed several times in phosphate-buffered saline containing 7% sucrose and 0.01% sodium azide and cryostat blocks were prepared. Dewaxed and cryostat sections were stained by either the indirect immunofluorescence method¹¹ or the peroxidase–antiperoxidase (PAP) technique¹². Sections were incubated for 16–20 h at 4 °C, with antisera diluted in phosphate-buffered saline.

For radioimmunoassay the two carotid bodies from each cat were pooled and extracted twice with 500 μ l of 0.1 M acetic acid or with 500 μ l of water and 500 μ l of 0.1 M formic acid. The extracts were assayed in serial dilutions for VIP, substance P- and enkephalin-like immunoreactivity. The nature of the enkephalin-like material was examined using both column chromatography (Fig. 3) and thin layer chromatography¹³ (TLC). The carotid bodies were composed of clusters of cells separated by connective tissue and a rich vascular network. Large vessels, nerves and occasional ganglion cells were observed in the periphery of the organ.

Following incubation with either Met- or Leu-enkephalin antisera, immunoreactive cells were observed in the carotid body (Fig. 1). These cells were arranged in groups corresponding to the type I cell clusters. Immunostaining of serial sections revealed that the antisera stained the same groups of cells. No immunoreactive nerve fibres were seen in sections incubated with enkephalin antisera and type I cells were not stained by either VIP or substance P antisera.

An extensive network of varicose nerve fibres containing VIP-like immunoreactivity was observed (Fig. 2), associated principally with blood vessels and enkephalin immunoreactive type I cell clusters. VIP-like material also occurred in nerve fibres around the carotid arteries and in ganglion cells in the periphery of the carotid body. Fine nerve fibres immunostained with substance P antiserum were also distributed throughout the organ. The staining of each antiserum was completely abolished by preabsorption of the antiserum with its respective antigen (Table 2).

Significant amounts of VIP- (71.7 ± 18.4 pmol per g ($\pm$ s.e.m.) $n = 5$), substance P- (54.6 ± 16.8 pmol per g) and enkephalin-like immunoreactivity (982.6 ± 203.8 pmol per g) were found in

the carotid body extracts. Examination of the extracts by TLC revealed that both Met- and Leu-enkephalin-like material was present, with between three and four times more Met- than Leu-enkephalin. On a BioGel column the enkephalin-like material eluted at the same position as ³H-labelled, synthetic Met-enkephalin (Fig. 3). In addition, a higher molecular weight immunoreactivity was detected which eluted with the void volume.

Our results reveal that, as in other organs⁸, the carotid body possesses an extensive network of nerve fibres containing substance P- and VIP-like immunoreactivity. The origin of these fibres is not entirely clear. Those in which VIP-like immunoreactivity occurs may, at least in part, arise locally, for scattered immunoreactive ganglion cells were seen in the periphery of the organ, whereas those containing substance P-like material may originate outside the carotid body. Substance P is reported to affect chemosensory activity in the cat carotid body¹⁴. However, the role of these peptides in this organ is unknown.

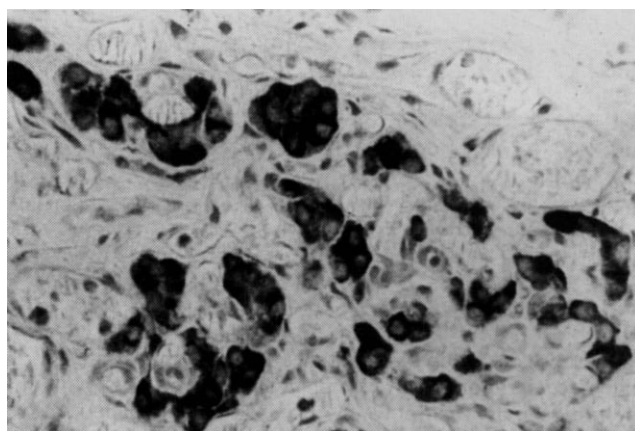


Fig. 1 Enkephalin-like immunoreactive type I cells in a 3- μ m thick section from a cat carotid body fixed in benzoquinone vapour ($\times 300$). Immunostained by the peroxidase–antiperoxidase technique with antiserum to Met-enkephalin (1:1,600 dilution).

Table 2 Characterisation of antisera for immunocytochemistry

Antiserum to:	REF	Titre		Absorption (0.1–40 nmol per ml of diluted antiserum)					
		IF	PAP	M-Enk	L-Enk	β -End	β -LPH	VIP	Sub P
Met-enkephalin	(497)	1:400	1:1,600	–(5)	–(40)	<(40)	<(40)	+	+
Met-enkephalin	(579)	1:200	1:1,000	–(5)	–(1)	<(20)	+	+	+
Leu-enkephalin	(493)	1:200	1:800	–(40)	–(5)	<(40)	+	+	+
Leu-enkephalin	(578)	1:200	1:1,000	–(5)	–(1)	–(20)	–(40)	+	+
VIP	(324)	1:300	1:2,000	+	+	+	+	–(0.1)	+
Substance P	(479)	1:500	1:4,000	+	+	+	+	+	–(5)

IF, Immunofluorescence technique¹¹; PAP, peroxidase–antiperoxidase technique¹²; REF, antiserum reference number. Controls included the application of antiserum preabsorbed with Met-enkephalin (M-Enk), Leu-enkephalin (L-Enk), β -endorphin (β -End), β -lipotropin (β -LPH), VIP or substance P (Sub P). +, No change in the intensity of the immunostaining; – (nmol ml^{–1}), minimum amount of peptide required to abolish the immunostaining; < (nmol ml^{–1}), amount of peptide required to reduce significantly the intensity of the immunostaining.

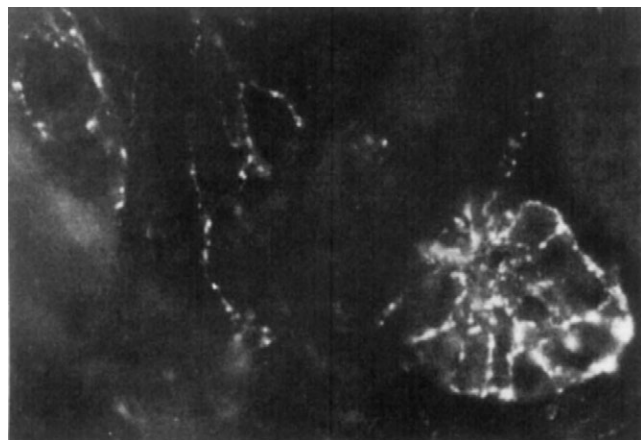


Fig. 2 Varicose nerve fibres containing VIP-like immunoreactivity, associated with a cluster of type I cells ($\times 415$). A 10- μm -thick cryostat section incubated with antiserum to VIP (1:300 dilution) and immunostained by the indirect immunofluorescence technique.

Several peptides are known to be stored with amines in a single cell^{6,15-17} or neurone¹⁸⁻²¹, and it has recently been reported that adrenal chromaffin cells^{22,23} and their tumours^{22,24} contain enkephalin-like peptides in addition to catecholamines. The finding of enkephalin-like material in carotid body type I cells provides another example of the co-existence of amines and peptides in the same cell.

Both β -endorphin and β -lipotropin contain the molecular sequence of Met-enkephalin^{25,26}; thus, the partial abolition of the immunostaining by these peptides (Table 2), particularly that of Met-enkephalin, is not unexpected and may indicate the presence of a larger precursor molecule. The larger molecular weight, immunoreactive substance found in the extracts may be similar to that previously reported to occur in sympathetic ganglia²⁷.

It is well known that opiates depress respiration²⁸, but little information is available regarding their role in the carotid body. Opiates affect the release of central and peripheral neurotransmitters²⁹⁻³⁴ and the catecholamine content of adrenal medullary cells^{32,35}. It has been proposed that the two predominant catecholamines in the carotid body, dopamine and noradrenaline, modulate chemoreceptor discharge³⁶⁻³⁸. Thus, the association of amines and peptides may be of importance for the chemoreceptor properties of the carotid body.

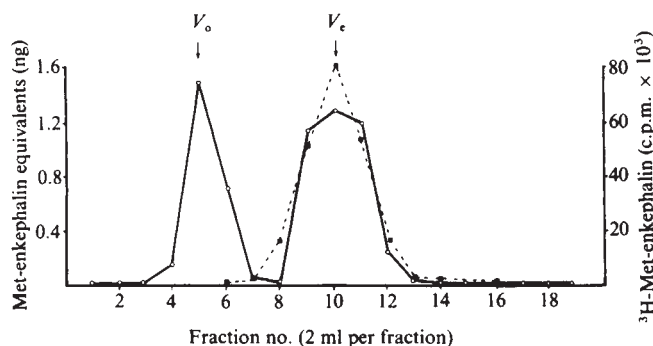


Fig. 3 Gel permeation chromatography profiles of a cat carotid body extract ($\circ$) and ^3H -labelled synthetic Met-enkephalin ($\blacksquare$). Pooled formic acid and water extracts (1 ml) were applied to a BioGel P2 column (10 cm $\times$ 1 cm) and eluted with a 100 mM sodium phosphate buffer (pH 7.4) containing 50 μM bacitracin, at a flow rate of 1 ml per 3 min 200,000 c.p.m. ^3H -Met-enkephalin was applied to the same column on a separate run, and detected in the effluent fractions by scintillation counting. V_0 , Void volume; V_e , enkephalin.

We thank Professor J. R. Garrett for supplying some of the carotid bodies used in this study. This work was supported by a grant from the Council for Tobacco Research.

Received 30 November 1979; accepted 28 January 1980.

1. Böck, P. & Gorgas, K. in *Chromaffin, Enterochromaffin and Related Cells* (eds Coupland, R. E. & Fujita, T.) 355-374 (Elsevier, Amsterdam, 1976).
2. Chiochio, S. R., King, M. P., Carballo, L. & Angelakos, E. T. *J. Histochem. Cytochem.* **19**, 621-626 (1971).
3. Kobayashi, Sh. *Arch. Histol. Jap.* **33**, 319-339 (1971).
4. Capella, C. & Solcia, E. *Virchows Arch. B* **7**, 37-53 (1971).
5. Kobayashi, S. in *Chemoreception in the Carotid body* (eds Acker, H. et al.) 221-227 (Springer, Berlin, 1977).
6. Pearse, A. G. E. *J. Histochem. Cytochem.* **17**, 303-313 (1969).
7. Polak, J. M. & Bloom, S. R. *Invest. Cell Path.* **1**, 301-326 (1978).
8. Polak, J. M. & Bloom, S. R. *Clin. Endocr. Metab.* **8**, 313-330 (1979).
9. Pearse, A. G. E. & Polak, J. M. *J. Histochem.* **7**, 179-186 (1975).
10. Bishop, A. E., Polak, J. M., Bloom, S. R. & Pearse, A. G. E. *J. Endocr.* **71**, 25P-26P (1978).
11. Coons, A. H., Leduc, E. H. & Connolly, J. J. *exp. Med.* **102**, 49-60 (1935).
12. Sternberger, L. in *Immunocytochemistry*, 142-161 (Prentice-Hall, Englewood Cliffs, 1974).
13. Hughes, J., Kosterlitz, H. W. & Smith, T. W. *Br. J. Pharmac.* **61**, 639-647 (1977).
14. McQueen, D. S. *J. Physiol., Lond.* **284**, 164P-165P (1978).
15. Aliments, J., Håkanson, R., Sundler, F. & Chang, K.-J. *Histochemistry* **56**, 187-196 (1978).
16. Heitz, Ph., Polak, J. M., Kasper, M., Timson, C. M. & Pearse, A. G. E. *Histochemistry* **50**, 319-325 (1979).
17. Owman, C. H., Håkanson, R. & Sundler, F. *Fedn Proc.* **32**, 1785-1791 (1973).
18. Chan-Palay, V., Jonsson, G. & Palay, S. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1582-1586 (1978).
19. Hökfelt, T., Elfvin, L.-G., Schultzberg, M., Goldstein, M. & Luft, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3587-3591 (1977).
20. Hökfelt, T. et al. *Neuroscience* **3**, 517-538 (1978).
21. Schultzberg, M. et al. *Neuroscience* **4**, 249-270 (1979).
22. Lundberg, T. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4079-4083 (1979).
23. Schultzberg, M. et al. *Neuroscience* **3**, 1169-1186 (1978).
24. Sullivan, S. N., Bloom, S. R. & Polak, J. M. *Lancet* **i**, 968-987 (1978).
25. Goldstein, A. *Science* **193**, 1081-1086 (1976).
26. Lazarus, L. H., Ling, N. & Guillemin, R. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2156-2159 (1977).
27. Di Giulio, A. M. et al. *Neuropharmacology* **17**, 989-992 (1978).
28. Florez, J. & Mediavilla, A. *Brain Res.* **138**, 385-590 (1977).
29. Eidelberg, E. *Prog. Neurobiol.* **6**, 81-102 (1976).
30. Hughes, J., Kosterlitz, H. W. & Leslie, F. M. *Br. J. Pharmac.* **53**, 371-381 (1975).
31. Jhamandas, K., Sawynok, J. & Sutak, M. *Nature* **269**, 433-434 (1977).
32. Maynert, E. W. & Klingman, G. J. *Pharmac. exp. Ther.* **135**, 285-295 (1962).
33. Taube, H. D., Borowski, E., Endo, T. & Starke, K. *Eur. J. Pharmac.* **38**, 377-380 (1976).
34. Pollard, H., Llorens-Cortez, C. & Schwartz, J. C. *Nature* **268**, 745-747 (1977).
35. Anderson, T. R. & Slotkin, T. A. *Biochem. Pharmac.* **24**, 671-679 (1975).
36. Mills, E., Smith, P. G., Slotkin, T. A. & Breeze, G. *Neuroscience* **3**, 1137-1146 (1978).
37. Nishi, K. in *Chemoreception in the Carotid Body* (eds Asker, H. et al.) (Springer, Berlin, 1977).
38. Zapata, P. J. *J. Physiol., Lond.* **244**, 235-251 (1975).

Selective inhibition of prostaglandin production in inflammatory exudates and gastric mucosa

B. J. R. Whittle, G. A. Higgs, K. E. Eakins, S. Moncada & J. R. Vane

Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

A major clinical problem encountered in the use of non-steroid anti-inflammatory drugs has been the high incidence of gastrointestinal irritation. The mechanisms underlying damage to the gastric mucosa evoked by these drugs are complex. The normal resistance of the gastric mucosa to the back-diffusion of gastric acid from the lumen into the mucosal tissue can be disrupted by topical administration of aspirin as well as several irritants such as ethanol, bile salts and detergents; this results in gastric mucosal damage¹. Such an action, however, cannot be the sole mechanism because many non-steroid anti-inflammatory drugs cause gastrointestinal damage when administered parenterally^{2,3}. It has been proposed that a reduction in prostaglandin biosynthesis by aspirin-like drugs⁴ in the gastric mucosa leads to a fall in local blood flow which, in turn, gives rise to areas of focal ischaemia ultimately developing into erosions⁵. However, a combination of the direct topical irritant actions and inhibition of prostaglandin formation can lead to a marked potentiation of gastric damage³. Such a situation is likely to be encountered in the clinical use of these compounds following

oral administration. We have investigated the relationship between gastric damage and inhibition of cyclooxygenase in the gastric mucosa following oral administration of anti-inflammatory compounds. For these studies, we have measured the *ex vivo* formation of prostacyclin (PGI₂), the major cyclooxygenase product in the rat gastric mucosa which, like other prostaglandins such as PGE₂, has gastro-protective actions⁶. In the inflammatory exudate induced by carrageenin, PGE₂ predominates and has pro-inflammatory actions⁷. We have, therefore, investigated whether selective inhibition of prostaglandin production, assayed as PGE₂-like activity, in the inflammatory exudate can be achieved *in vivo*. We report here that it can.

Previous studies have shown that cyclooxygenase inhibition *in vivo* prevents subsequent prostacyclin generation by the rat gastric mucosa *ex vivo*⁸. Strips of gastric mucosa (0.2–0.4 g) freed from underlying muscle were rapidly washed, chopped and incubated in buffer (50 mM Tris, pH 8.4, 1 ml) by vortex mixing (1 min, 22 °C). After centrifugation in a fixed-speed Eppendorf bench centrifuge (15 s at 9,000g) the supernatant was immediately tested for its ability to inhibit human platelet aggregation. This prostacyclin-like activity, which was characterised as described before⁸ and by the use of an antiserum which binds and inactivates prostacyclin⁹, was assayed against authentic prostacyclin.

Inflammatory exudates in male rats (200 g body weight) were induced and collected by the subcutaneous implantation of polyester sponges impregnated with carrageenin (20 mg ml⁻¹ in sterile saline). After 24 h, the sponges were removed, immersed in heparinised saline (5 ml) and squeezed until dry. The pros-

taglandin-like activity in acid-lipid extracts of these exudates was determined by bioassay on the rat superfused stomach strip, treated with a mixture of antagonists, and was expressed as PGE₂ equivalents¹⁰.

Erosion formation in the gastric mucosa was determined as described previously³. Food was removed from the individually caged rats immediately after initial drug administration. The animals were killed after 24 h and the incidence and severity of the erosions, which formed only in the secretory mucosa, was assessed. The compounds were suspended in Celacol solution (0.25% w/v carboxymethylcellulose) and the suspension adjusted to pH 5–6. Drugs were administered orally (0.1 ml per 100 g body weight), three times during the 24 h period, at the time of sponge implantation, 6 h later and again 3 h before removal of sponge or assessment of mucosal erosion formation and prostacyclin production.

The doses of the compounds given (see Fig. 1 legend) are all in the range causing 50–70% inhibition of inflammation, as determined by reduction in the oedema induced 3 h after the subplantar injection of carrageenin in the hind paw of rats^{11,12}. The dose causing 50% reduction of rat paw oedema (ED₅₀) was (in mg per kg) for aspirin, 130; sodium salicylate, 140; indomethacin, 2.7; flurbiprofen 0.4; naproxen, 5. The recently described dual inhibitor of cyclooxygenase and lipoxygenase, BW755C (3-amino-1[(*m*-trifluoromethyl)-phenyl]-2-pyrazoline)¹² was administered in doses greater than those required for the anti-inflammatory actions (ED₅₀ 11 mg per kg).

The clinically used compounds all reduced cyclooxygenase products in both the sponge and the gastric mucosa in these anti-inflammatory doses (Fig. 1B). Both aspirin and flurbiprofen seemed more active on cyclooxygenase in the gastric mucosa than in the inflammatory exudate. In contrast, sodium salicylate and BW755C, in doses causing a significant reduction in prostaglandins in the sponge, failed to inhibit mucosal cyclooxygenase activity. In further studies, the compounds were administered by a single oral dose and prostacyclin generation by the gastric mucosa was determined 3 h later. Sodium salicylate (200 mg per kg) and BW755C (100 mg per kg) again failed to prevent mucosal cyclooxygenase activity, whereas aspirin (50–200 mg per kg) and indomethacin (1.25–10 mg per kg) were potent inhibitors.

The activity of these compounds on gastric mucosal cyclooxygenase activity 3 h after subcutaneous injection was also investigated. Sodium salicylate (50–400 mg per kg) and BW755C (25–150 mg per kg) caused no significant inhibition of *ex vivo* prostacyclin formation, whereas aspirin (50–200 mg per kg), indomethacin (1.25–10 mg per kg), naproxen (2.5–20 mg per kg) and flurbiprofen (0.1–1 mg per kg) all caused dose-dependent reductions in prostacyclin production. This indicates that the potent inhibition of mucosal cyclooxygenase by these latter compounds is not simply a consequence of selective accumulation in the gastric mucosal tissue following administration of local high concentrations of these agents.

To investigate whether the failure to prevent prostacyclin formation was due to washout or dilution of the drug from the mucosa during the incubation procedures, gastric mucosal tissue from rats pretreated with sodium salicylate (200 mg per kg) was washed, chopped and incubated in buffer solutions containing sodium salicylate (10–100 µg ml⁻¹). Again, no inhibition of prostacyclin production could be detected. It has been suggested that the inhibition of prostacyclin production in rat aortic rings *in vitro* by indomethacin pretreatment can be reversed by washing and incubation of the tissue¹³. However, with our procedure and using gastric mucosal tissue, we have shown that pretreatment of rats with aspirin, indomethacin, naproxen and flurbiprofen always led to a potent inhibition of prostacyclin production *ex vivo*, suggesting little tissue washout or irreversible cyclooxygenase inhibition with these drugs.

The formation of gastric erosions following oral administration of those agents in anti-inflammatory doses, three times over 24 h, is shown in Fig. 1A, expressed in terms of an erosion

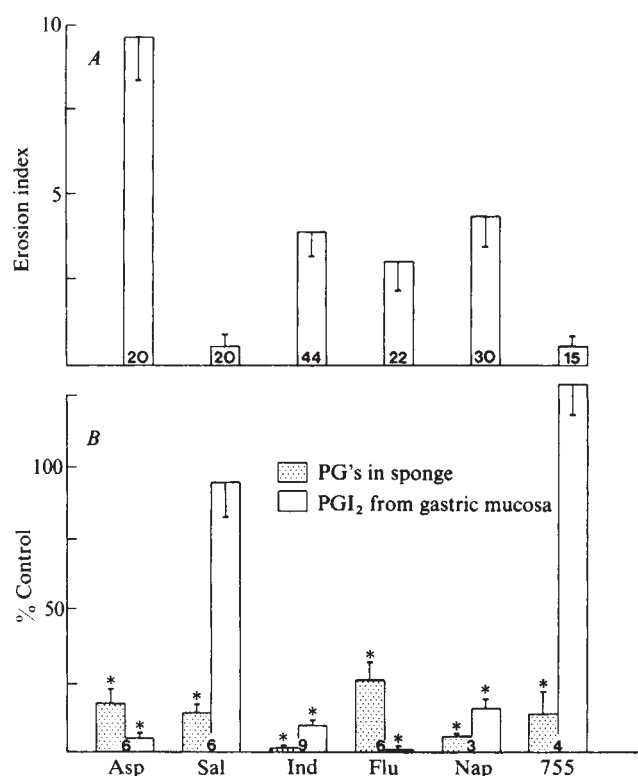


Fig. 1 Effect of oral administration of non-steroid anti-inflammatory agents on the induction of gastric erosions in rats (A), and prostaglandin E₂-like activity in the sponge inflammatory exudate and the formation of prostacyclin (PGI₂) by the gastric mucosa *ex vivo* (B). The results are expressed as the mean ± s.e.m. of *n* values. The changes in prostaglandin levels are given as % change compared with control levels, where an asterisk indicates significance (*P* < 0.01). The degree of gastric damage is expressed in terms of an erosion index. Drugs were administered orally three times over 24 h in the following doses (mg per kg): aspirin (200); sodium salicylate (200); indomethacin (4); flurbiprofen (0.5); naproxen (5); BW755c (100).

index. Indomethacin, flurbiprofen and naproxen induced a comparable degree of gastric erosion formation which paralleled their efficacy in inhibiting mucosal prostacyclin production in the doses used. Aspirin, however, was more active in inducing gastric damage in doses causing a similar mucosal cyclooxygenase inhibition to the other compounds. This may reflect the synergistic interaction³ between the topical irritation known to occur with aspirin^{1,2} and the inhibition of mucosal cyclooxygenase.

Interestingly, sodium salicylate, which failed to inhibit mucosal prostacyclin production in anti-inflammatory doses, caused little gastric damage. This finding is similar to observations in man that sodium salicylate produced far less gastrointestinal irritation than aspirin, as determined by faecal blood loss^{2,14}. Thus, topical irritation alone may not lead to pronounced gastric bleeding and erosion formation. Previously, it has been shown that systemically administered salicylic acid failed to prevent thromboxane A₂ formation in rat platelet-rich plasma, a situation in which washout of an active drug could not occur¹⁵. Furthermore, in our own studies on prostacyclin production *ex vivo* by rat isolated aortic rings, prior systemic administration of sodium salicylate (200 mg per kg, subcutaneously) failed to inhibit the vessel wall cyclooxygenase. This suggests that cyclooxygenase activity at the site of inflammation may be uniquely sensitive to inhibition by sodium salicylate and perhaps other drugs. An alternative, although unlikely explanation, would be that sodium salicylate could selectively inhibit PGE₂ but not prostacyclin formation, by an action beyond the cyclooxygenase, perhaps by affecting the subsequent transformation of the endoperoxide intermediate.

The dual cyclooxygenase and lipoxygenase inhibitor, BW755C, which also failed to inhibit mucosal cyclooxygenase activity, similarly caused little gastric damage. This again gives support to a relationship between the production of gastric erosions and the inhibition of mucosal prostacyclin formation by anti-inflammatory drugs. The doses of BW755C used in this study were 5–10 times greater than the anti-inflammatory doses required to reduce carrageenin foot paw oedema and white cell migration into sponges¹², suggesting a good therapeutic ratio for this experimental compound. Further, the dual inhibition of the two major pathways of arachidonate metabolism is likely to be a beneficial property in the control of leukocyte migration and inflammation¹².

Although cyclooxygenase enzyme preparations from tissues such as spleen, kidney and brain have shown differential sensitivities to some compounds *in vitro*¹⁶, the present findings clearly demonstrate that some non-steroid anti-inflammatory drugs can selectively inhibit prostaglandin biosynthesis in different tissues *in vivo*. The development of anti-inflammatory compounds which fail to inhibit cyclooxygenase in the gastric mucosa and lack topical irritancy seems to be a rational approach to obtaining clinically well-tolerated drugs.

We thank Mr K. Mugridge and Ms J. Walker for technical assistance and Drs. C. V. Denyer and F. C. Copp for preparing BW755C.

Lentil lectin effectively induces allotransplantation tolerance in mice

Ivan Hilgert, Václav Hořejší, Pavla Angelisová & Hana Křištofová

Institute of Molecular Genetics, Czechoslovak Academy of Sciences, Prague 4, Czechoslovakia

The interaction of lectins with carbohydrate receptors on the plasma membrane of eukaryotic cells results in a wide variety of biological effects¹. One effect which has been extensively studied is the stimulation of lymphocytes to blastogenesis and proliferation. High doses of concanavalin A (Con A) and phytohaemagglutinin (PHA) have been shown to activate *in vitro* regulatory cells capable of suppressing proliferation of other cells in the culture^{2,3}. However, Con A and PHA treatment were used effectively to prolong the survival of skin and heart allografts^{4–8} before it was recognised that some lectins have an activatory effect on suppressor cells *in vitro*. A possible explanation of these tolerogenic effects is the activation of specific suppressor cells. In this letter we have compared systematically various lectins differing in their carbohydrate specificity and mitogenicity in relation to their ability to induce prolonged skin allograft survival in mice with the aim of selecting the most effective lectin treatment schedule. Some preliminary results of this study have been mentioned in a recent review⁹.

All lectins used were purified by affinity chromatography and were devoid of any electrophoretically detectable contaminants. Con A, pea lectin (PsA), lentil lectin (LcA) and sweet pea (*Lathyrus odoratus*) lectin (LoA) were prepared by affinity chromatography on Sephadex G-100 (refs 10–12), soybean lectin (SBA) by affinity chromatography on *N*-acetyl-D-galactosaminyl polyacrylamide gel¹³, *Erythrina indica* (EiA) lectin on α -D-galactosyl polyacrylamide gel¹⁴, wheat germ lectin (WGA) on *N*-acetyl- α -D-glucosaminyl polyacrylamide gel and scarlet runner (*Phaseolus cocineus*) seed lectin (PcA) by affinity chromatography on fetuin-Sepharose. Unless otherwise stated, mouse strain combination B10.D2 $\times$ B10.D2(M504) (abbreviated to M504) was used; the histocompatibility difference between these strains lies in their H-2D regions^{15,16}. In other experiments another strain combination was used (B10.A $\times$ C57BL10ScSn) (abbreviated to B10) these strains differing in all regions of the H-2 locus. To study the specificity of tolerance induced by the lectin treatment and B10.D2 skin graft in M504 recipients, third-party grafts from BALB/c mice were used. This strain is identical with the original donor strain (B10.D2) at the H-2 locus but differing at the multiple non-H-2 loci. In all cases, both recipients and donors were females.

Comparisons were made using three different time schedules of LcA injection relative to time of skin allografting (Fig. 1). In schedules where all 10 doses of the lectin were given before grafting or 5 doses were given before and 5 after grafting, allograft survival was not essentially affected. In contrast, 10 doses (each 250 μ g) given after grafting made 2 out of 6 animals long-term tolerant (skin graft survival >100 days). Further improvement of graft survival time could be achieved by increasing the lectin dose to 500 μ g and total number of doses given to 20. In these conditions allografts of all LcA-treated recipients survived for more than 100 days (Fig. 2). These long-term surviving grafts did not show any macroscopically visible signs of rejection.

The effect of different lectins was compared using this optimum schedule (20 doses of 500 μ g, all given intraperitoneally (i.p.) after transplantation) (Fig. 2). (The doses of Con A and WGA, however, were only 250 μ g because of their toxicity at higher doses.) WGA, SBA and EiA did not induce long-term

Received 12 December 1979; accepted 29 January 1980.

1. Davenport, H. W. *Gastroenterology* **46**, 245–253 (1964).
2. Grossman, M. I., Matsumoto, K. K. & Lichter, R. J. *Gastroenterology* **40**, 383–388 (1961).
3. Whittle, B. J. R. *Br. J. Pharmac.* **60**, 455–460 (1977).
4. Vane, J. R. *Nature new Biol.* **231**, 232–233 (1971).
5. Main, I. H. M. & Whittle, B. J. R. *Br. J. Pharmac.* **53**, 217–224 (1975).
6. Whittle, B. J. R., Boughton-Smith, N. K., Moncada, S. & Vane, J. R. *Prostaglandins* **15**, 955–967 (1978).
7. Higgs, G. A. & Salmon, J. A. *Prostaglandins* **17**, 737–746 (1979).
8. Whittle, B. J. R. *J. Pharm. Pharmac.* **30**, 467–468 (1978).
9. Bunting, S., Moncada, S., Reed, P., Salmon, J. A. & Vane, J. R. *Prostaglandins* **15**, 565–573 (1978).
10. Higgs, G. A., Harvey, E. A., Ferreira, S. H. & Vane, J. R. in *Advances in Prostaglandin and Thromboxane Research* Vol. 1 (eds Samuelsson, B. & Paoletti, R.) 105–110 (Raven, New York, 1976).
11. Winter, C. A., Risley, E. A. & Nuss, G. W. *Proc. Soc. exp. Biol. Med.* **111**, 544–547 (1962).
12. Higgs, G. A., Flower, R. J. & Vane, J. R. *Biochem. Pharmac.* **28**, 1959–1961 (1979).
13. Hornstra, G., Haddeman, E. & Don, J. A. *Nature* **279**, 66–68 (1979).
14. Leonards, J. R. & Levy, G. *Clin. Pharmac. Ther.* **14**, 62–66 (1973).
15. Vargaftig, B. B. *J. Pharm. Pharmac.* **30**, 101–104 (1978).
16. Flower, R. J. & Vane, J. R. *Prostaglandin Synthetase Inhibitors* (eds Robinson, H. J. & Vane, J. R.) 9–18 (Raven, New York, 1974).

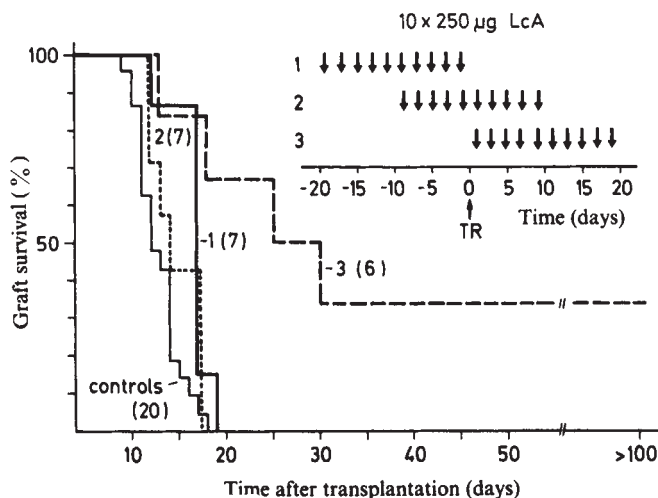


Fig. 1 Survival of B10.D2 grafts in M504 recipients treated with 10 doses of LcA (250 µg, i.p.). Comparison of three different time schedules of LcA administration. Inset shows the schedule of treatment (intervals between the doses were 48 h). Test transplantation (TR) is always on day 0. Skin grafts prepared from tail skin ~0.5 cm² in size were transplanted onto the backs of the tested recipients²⁵. Grafts with irreversible necrotic changes over their whole area were considered rejected.

skin allograft survival (date not shown), Con A, PcA and LoA induced tolerance only occasionally, whereas PsA induced long-term tolerance in about 50% of recipients and LcA in 100% of recipients. The LcA treatment did not seem to have any observable harmful side effects on the animals.

The efficiency of the most effective lectin, LcA, could be still increased by administering some doses intravenously (i.v.) (1st, 3rd, 5th and 7th dose given i.v., all others i.p.); in this situation, 20 doses of only 250 µg each induced long-term allograft survival in 90–100% of recipients. Long-term tolerance in 90–100% of recipients was induced also if the most effective lectin treatment started on the third day after transplantation; when it started 5 days after skin grafting, long-term tolerance was still induced in 50% of the cases.

Transplantation of a second B10.D2 allograft (no treatment) into recipients surviving with the first graft after 100 days as a result of LcA treatment led almost always to rejection of the

second graft in 12–20 days; the primary graft remained unaffected. This indicates either a rather unstable nature of the lectin-induced tolerance even 100 days after primary graft transplantation, or the occurrence of some form of protection of the first graft in the host. Although unexpected, this result is not unprecedented; similar effects were observed in the case of neonatally-induced tolerance (G. A. Voisin, personal communication) and in the case of liver extract-induced tolerance¹⁷, where the stability of tolerance was shown to increase progressively with time elapsed after transplantation (from 50th to 150th day). Studies on the possible stabilisation of lectin-induced tolerance with time and by increased lectin dose are now in progress in our laboratory.

Third-party test allografts (BALB/c) transplanted into recipients bearing a B10.D2 graft for >100 days resulted in a rapid rejection of the BALB/c graft (11–14 days), usually accompanied by a delayed rejection of the primary B10.D2 graft (in 15–30 days).

Therefore, the mechanism of this type of tolerance is not yet clear, but the lectin certainly does not act through nonspecific damage or suppression of immunological reactivity of the recipient as evidenced by the rapid rejection of third-party grafts. It is of interest that the lectin treatment is remarkably efficient even when it is started several days after transplantation; this indicates that the lectin might be interacting with the antigen-sensitised lymphocytes.

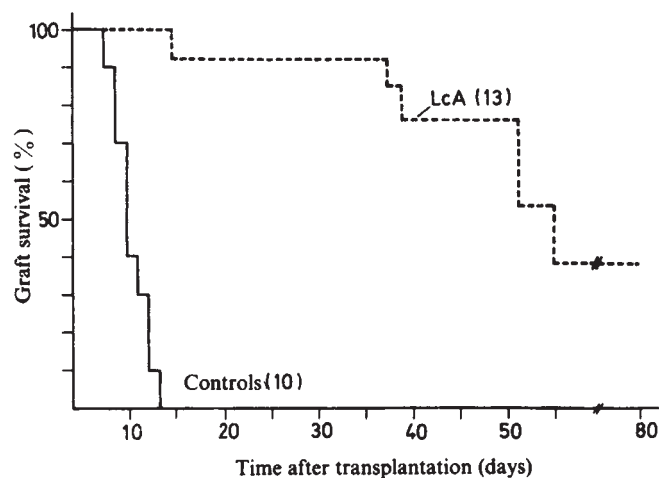


Fig. 3 Survival of B10.A grafts in B10 recipients treated with LcA. The lectin (1 mg in 0.5 ml phosphate buffered saline) was given i.p. daily starting with the day of transplantation (day 0) until rejection; the doses on days 1, 3, 5, 7 and 9 were given i.v. For other experimental details see Fig. 1 legend.

All lectins with detectable tolerogenic activity (except PcA) have similar carbohydrate specificity (α -D-Man, α -D-G1c) and are structurally fairly similar (a Con A-like group of lectins)¹⁸. Lectins with similar carbohydrate specificity and probably also similar structure are present in seeds of many other plants belonging to the Fabaceae family¹⁹; it is quite possible that some of these poorly studied lectins might be even more effective tolerance inducers than LcA.

The primary target of action of LcA and similar lectins is probably a specific carbohydrate receptor or class of receptors on the cell surface. However, the plant lectins might also act by mimicking the action of (or competing with) some endogenous lectin-like regulatory substances (for example, lymphokines).

The LcA treatment of recipients compares well with other treatments used for induction of adult transplantation tolerance (spleen extract alone²⁰ or in combination with hydrocortisone acetate²¹, antilymphocyte serum²²) at least in the mouse strain

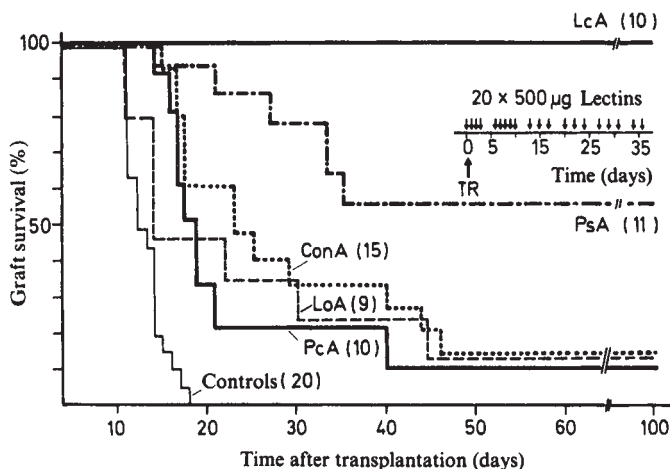


Fig. 2 Survival of B10.D2 grafts in M504 recipients treated with 20 x 500 µg i.p. doses of different lectins. The dose of Con A was 250 µg due to its toxicity in higher doses. Inset time schedule of lectin treatment.

combination tested (B10.D2×M504); unlike LcA, none of these treatments was able to induce long-term (permanent) survival of skin grafts in 100% of the recipients. Other highly effective ways of induction of transplantation tolerance in mice, for example combined treatment with several immunosuppressive drugs²³ or cyclosporin A treatment²⁴, were demonstrated in other strain combinations and thus direct comparison with LcA effects is not possible at present. Our recent results indicate that LcA is highly efficient in prolonging allograft survival and in inducing tolerance in a strain combination with a strong H-2 barrier (B10.A×B10) (Fig. 3).

In conclusion, LcA is possibly one of the most effective substances known so far (in combination with skin graft as a source of specific antigens) for inducing adult transplantation tolerance in mice. Experiments are now in progress to determine what maximal histocompatibility barrier can be overcome by LcA treatment alone or in combination with other immunosuppressive drugs, whether LcA is also effective in other animal species, the details of induction mechanism and maintenance of this type of tolerance.

LoA was provided by Dr M. Tichá.

Received 6 November 1979; accepted 10 January 1980.

- Nicolson, G. L., *Int. Rev. Cytol.* **39**, 89–100 (1974).
- Peary, D. L. & Pierce, C. W. *J. exp. Med.* **140**, 356–369 (1974).
- Rich, R. R. & Rich, S. S. *J. Immun.* **114**, 1112–1115 (1975).
- Markley, K., Thornton, S., Smallman, E. & Markley, P. *Transplantation* **8**, 258–264 (1969).
- Markowitz, M., Person, D. A., Gitnick, G. E. & Ritts, R. E. *Science* **163**, 476 (1969).
- Stefani, S. S. & Moore, C. D. *J. Immun.* **104**, 780–784 (1970).
- Rosenau, W., Haber, J. & Goldberg, M. *Transplantation* **13**, 624–626 (1972).
- Weil, III R. *et al. Transplantation* **17**, 600–604 (1974).
- Hilgert, I. *Immun. Rev.* **46**, 27–53 (1979).
- Agrawal, B. B. L. & Goldstein, I. J. *Biochim. biophys. Acta* **147**, 262–271 (1967).
- Entlicher, G., Košťř, J. V. & Kocourek, J. *Biochim. biophys. Acta* **221**, 272–281 (1970).
- Tichá, M., Entlicher, G., Košťř, J. V. & Kocourek, J. *Biochim. biophys. Acta* **221**, 282–289 (1970).
- Hofejší, V. & Kocourek, J. *Biochim. biophys. Acta* **538**, 299–315 (1978).
- Hofejší, V., Tichá, M., Novotný, J. & Kocourek, J. *Biochim. biophys. Acta* (in the press).
- Dishkant, I. P., Vedernikov, A. A. & Egorov, I. K. *Genetika* **9**, 83–90 (1973).
- Klein, J. *Adv. Immun.* **26**, 55–146 (1978).
- Hilgert, I., Vlachov, K. & Viklický, V. *Transplantation* **18**, 471–479 (1974).
- Goldstein, I. J. & Hayes, C. E. *Adv. Carbohydr. Chem. Biochem.* **35**, 127–340 (1978).
- Tobiška, J. *Die Phytinmagglutinine* (Akademie, Berlin, 1964).
- Hilgert, I., Křištofová, H. & Rieger, M. *Folia biol., Praha* **21**, 409–413 (1975).
- Rieger, M., Günther, J., Křištofová, H. & Hilgert, I. *Folia biol., Praha* **24**, 146–161 (1978).
- Rieger, M., Hilgert, I., Křištofová, H. & Vlachov, K. *Folia biol., Praha* **25**, 220–230 (1979).
- Brent, L. & Opara, S. C. *Transplantation* **27**, 120–126 (1979).
- Borel, J. F., Feuer, C., Gubler, H. U. & Stähelin, H. *Agents and Actions* **6/4**, 468–475 (1976).
- Billingham, R. E. & Medawar, P. B. *J. exp. Biol.* **28**, 385–402 (1951).

The human 'T' genetic region of the HLA linkage group is a polymorphism detected on lectin-activated lymphocytes

Ephraim Gazit

Tissue Typing Laboratory, Chaim-Sheba Medical Center, Tel-Hashomer Hospital, Israel

Cox Terhorst & Edmond J. Yunis

Divisions of Tumor Immunology and Immunogenetics, Sidney Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115

The major histocompatibility complex of the mouse, H-2, comprises 12 loci which are grouped into regions and subregions¹. Present data indicate that to the 'right' of the H-2D there are nine loci clustered into one region, the 'T region'. These loci are designated Qa-1, Qa-2, Qa-3, Qa-4, Qa-5, H-2T, H-31, H-32, TLa (ref. 2). The loci in the 'T region' may, in fact, be part of the H-2 complex. This notion is supported by the fact that gene products of the TLa, Qa-2 and Qa-3 loci share some molecular properties with H-2 products: a two-polypeptide chain complex with a common determinant of molecular weight 12,000 β_2 -microglobulin and a variable heavy chain of MW 45,000 (ref. 3). Recent reports suggested the possible existence of human analogues to the gene products of the murine 'T region'. Antigenic structures, similar but not identical with HLA-A, -B and -C antigens, were recently isolated from the human lymphoblastoid T-cell line MOLT4 (refs 4, 5). These reports presented evidence for the existence of a β_2 -microglobulin or a β_2 -microglobulin-like molecule, Bt, associated with a polypeptide chain of MW ~45,000. However, it remains to be determined whether these structures are equivalent to the mouse Qa-1, Qa-2 or TLa gene products. We now report the identification of human alloantisera that react with cell-surface determinants expressed on mitogen-activated T lymphocytes but not expressed on resting T cells. Antigen blocking and modulation techniques were used. The alloantigens identified (HT) are different from the classical HLA-A, -B, -C and -DR antigens but are associated with β_2 -microglobulin. These determinants are encoded by genes in the HLA-A region and may represent the human counterpart of the TLa region determinants of the mouse¹⁻³.

We have screened 200 sera obtained from women with a history of multiple pregnancies. The sera were exhaustively absorbed with platelets and were shown to be non-reactive with HLA-A, -B and -C antigens carried by the resting peripheral blood lymphocytes from two different individuals: D.W., whose HLA phenotype is A1, A3, B8, B18, Cw2, DRw3, DRw5; and E.G., whose HLA phenotype is A2, A3, B14, B37, Cw6, DRw1 and DRw5. Lectin (phytohaemagglutinin (PHA) or concanavalin A (Con A))-activated lymphocytes obtained from these two individuals were used to screen the platelet-absorbed pregnancy sera. Two antisera, SF48 and K, were found to react exclusively with the activated lymphocytes by a cytotoxicity score of 40–60%.

The sera were subsequently absorbed with resting lymphocytes from either D.W. or E.G. by a tray microabsorption technique (200×10^6 lymphocytes per ml) and the sera were found to retain their reactivity with mitogen-activated lymphocytes only (Table 1). The sera were then tested in a random panel of activated lymphocytes and gave 11/23 (47.8%) and 8/23 (34.8%) positive reactions with SF48 and K, respectively. The 2×2 tables between the HLA assignment of the panel and the cytotoxic reactions of these two sera are given in Table 2. Significant associations were found for reactivities of serum SF48 with HLA-A3 ($P = 1.77 \times 10^{-6}$) and serum K with HLA-A1 ($P = 1.24 \times 10^{-3}$). Having serologically ruled out the existence of antibodies against HLA antigens, we tested these sera in a selected panel of HLA-A1 or HLA-A3 phenotypes and demonstrated that lectin-activated lymphocytes in three out of five HLA-A1 individuals reacted with K, and five out of seven HLA-A3 individuals reacted with SF48 (Table 3). In addition, three T-lymphoblastoid cell lines were tested with SF48 and K antisera: HSB-2 (HLA-A1, A2, B17, Bw44), RPMI-8402 (HLA-A1, A29, B18, B16) and MOLT-4 (HLA-A1₂A10). Only MOLT-4 was able to absorb both the SF48 and K reactivity directed against D.W. and E.G. test cells, but did not remove the reactivity of HLA-A3 antiserum. MOLT-4, in contrast to the activated lymphocytes (see Table 1), reacted with 100% cytotoxicity, indicating that all MOLT-4 cells carry the antigen(s) detected by SF48 and K. It is of interest that the MOLT-4 cells have recently been shown to contain antigenic structures different from HLA-A and -B but associated with β_2 -microglobulin^{4,5}. These experiments, *per se*, exclude the possibility that the K and SF48 antigenic determinants represent a presentation of HLA-A, -B or -C antigens that are unique to lectin-stimulated lymphocytes.

To examine the remote possibility that HLA antibodies were not removed from SF48 and K sera after sequential absorptions

Table 1 Serological characterisation of PHA- and Con A-activated lymphocytes

Donor	Cell tested	Pretreatment	Microcytotoxicity scores with			
			Alloantisera	K	Heteroantisera	
			SF48		goat anti- β_2 m	rabbit anti p29,34(C91)
D.W.	E ⁺ lymphocytes	—	1	1	8	1
	E ⁻ lymphocytes	—	1	1	8	8
	PHA-activated lymphocytes	—	6	6	8	4
	Con A-activated lymphocytes	—	6	6	8	4
	Absorption of sera on:					
	PHA-activated lymphocytes	lymphocytes	6	6	1	1
	PHA-activated lymphocytes	E ⁺ lymphocytes	6	6	1	4
	PHA-activated lymphocytes	E ⁻ lymphocytes	6	6	1	1
	Blocking with:					
	PHA-activated lymphocytes	Turkey anti-p29,34	6	6	8	1
E.G.	PHA-activated lymphocytes	Turkey anti- β_2 m	2	1	1	4
	Con A-activated lymphocytes	Turkey anti-p29,34	6	6	8	1
	Con A-activated lymphocytes	Turkey anti- β_2 m	2	1	1	4
	E ⁺ lymphocytes	—	1	1	8	1
	E ⁻ lymphocytes	—	1	1	8	8
	PHA-activated lymphocytes	—	4	1	8	4
	Con A-activated lymphocytes	—	4	1	8	4
	Absorption of sera on:					
	PHA-activated lymphocytes	lymphocytes	4	1	1	1
	PHA-activated lymphocytes	E ⁺ lymphocytes	4	1	1	4
	PHA-activated lymphocytes	E ⁻ lymphocytes	4	1	1	1
	Blocking with:					
	PHA-activated lymphocytes	Turkey anti-p29,34	4	1	8	1
	PHA-activated lymphocytes	Turkey anti- β_2 m	1	1	1	4
	Con A-activated lymphocytes	Turkey anti-p29,34	4	1	8	1
	Con A-activated lymphocytes	Turkey anti- β_2 m	1	1	1	4

The platelet-absorbed pregnancy sera SF48 and K were tested in a microcytotoxicity assay, using fluorescein diacetate¹². The numbers are estimates obtained in the two-stage microlymphocytotoxicity assay: 1, 2, 4, 6 and 8 denote 0%, 10–25%, 25–50%, 50–75% and 75–100% cytotoxicity above background, respectively. Peripheral blood lymphocytes were separated on a Ficoll-Hypaque density gradient, washed three times and cultured at 37 °C in RPMI-1640 containing 20% pooled human serum, penicillin, streptomycin and L-glutamine. The cells were cultured for 5 d at a cell density of 1×10^5 per ml with either 5 μ g Con A per ml or 1.201 PHA per ml. E⁺ and E⁻ lymphocytes were obtained after rosetting with neuraminidase-treated sheep red blood cells¹³. Absorptions were carried out with 50,000 cells per μ l of serum in the case of E⁺ and E⁻ cells and with 200,000 cells per μ l of serum when whole lymphocytes were used. Blocking was done as previously published¹⁴. Turkey and goat-anti- β_2 -microglobulin (β_2 m) sera were prepared by multiple injections (in complete Freund's adjuvant) of β_2 -microglobulin purified from the urine of kidney transplant patients¹⁵. Turkey and rabbit (C91) anti-p29,34 (anti-Ia) sera were prepared in a similar way. The p29,34 antigens were isolated from a highly purified plasma membrane preparation¹⁶ of the B-lymphoblastoid cell line JY after solubilisation in deoxycholate by lentil lectin affinity chromatography and anti- β_2 -microglobulin immunoabsorbent column chromatography as described earlier^{17,18}. All sera were heat inactivated before use.

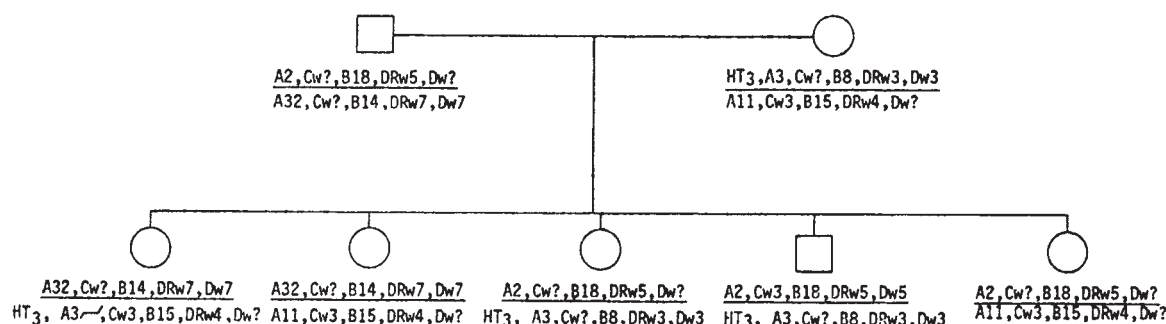


Fig. 1 A family (family F) using a microcytotoxicity assay with serum SF48. The first sibling on the left is an HLA-A/B recombinant. This family was recently discovered and has been thoroughly studied in the 8th International Histocompatibility Workshop.

with pooled platelets and resting lymphocytes from the same donor of the activated lymphocytes, the following experiment was carried out. Activated lymphocytes from individuals who were: (1) HLA-A3⁺, SF48⁺, (2) HLA-A3⁺, SF48⁻, and (3) HLA-A3⁻, SF48⁻, were used in cross-absorption experiments, in which HLA alloantisera defining HLA-A3 and SF48 were tested. The results obtained indicated that HLA-A3⁺, SF48⁺

cells absorbed all activity from both types of antisera, HLA-A3⁺, SF48⁻ cells absorbed the HLA only and HLA-A3⁻, SF48⁻ cells absorbed neither of them (Table 4).

Second, antigenic modulation experiments were carried out using a rabbit anti-HLA heavy chain serum, C41. For the preparation of the rabbit anti-HLA heavy chain antiserum (C41), the larger polypeptide chain of papain-solubilised HLA-

Table 2 2×2 tables of association between HLA assignment of the panel and SF48 and K reactivity

SF48	HLA-A3		K	HLA-A1	
	+	-		+	-
	+11	1		+	4
	- 0	11		- 0	15
$P(\text{Fisher}) = 1.77 \times 10^{-6}$			$P(\text{Fisher}) = 1.24 \times 10^{-3}$		

HLA assignment of the screening panel compared with the reactivity to SF48 and K. The panel consisted of unrelated randomly chosen individuals ($n = 23$). The P value reflects the probability of a random association between the two variables under study.

A2 antigen was isolated from the B-lymphoblastoid cell line RPMI-4265 as described elsewhere⁶. The resulting antibody had a cytotoxicity titre of 1:100 using lymphoblastoid cell lines JY and 1:10 using resting peripheral lymphocytes. This antibody also reacted with Daudi cells which agrees with studies showing that Daudi cells express the heavy chain of HLA antigens on their surface⁷. In addition, two monoclonal HLA heavy chain antibodies (w6/32 and w34/28)⁸ were obtained commercially (from Accurate Chemical Corporation). Both resting and mitogen-activated lymphocytes were incubated with varying dilutions of these antisera for periods extending to 6 h. The lymphocytes were then washed and retested with the three anti-HLA heavy chain sera or with alloantibodies. Results showed that prior incubation of the activated lymphocyte with HLA heavy chain antibody, in the absence of complement, did not suppress HLA heavy chain determinants on retesting in a subsequent cytotoxicity assay. Indeed, all three HLA heavy chain antibodies, in addition to HLA-A1, HLA-A3, SF48 and K antisera, were still capable of lysing the test cells. However, antigenic modulation of activated lymphocytes from E.G. (A2, A3, B14, B37, Cw6) with HLA-A3 antibody (kras) eliminated subsequent lysis mediated by HLA-A3 and complement, but not lysis by SF48 and complement. Following antigenic modulation with HLA-A3 antibody in control experiments, subsequent lysis was positive with HLA-A2 and HLA-B14 antisera but negative with HLA-A1 antiserum and complement. These results indicated that SF48 and HLA-A3 membrane antigens are expressed independently on the cell surface of mitogen-activated lymphocytes.

As Ia antigens are expressed on activated T cells⁹⁻¹¹, we designed experiments to determine whether or not the reactivity in SF48 and K alloantisera is due to antibody directed against DRw antigens. Blocking experiments were carried out by pretreating Con A- and PHA-activated T cells with a turkey

Table 3 Comparison of HLA and DR assignments to those of SF48 and K reactivities in a selected screening panel

Individuals	HLA	DRw	SF48	K
T.R.	A2, A3, B17, B18	7	+	+
K.F.	A1, A1, B8, B8	3,3	+	+
B.M.	A3, Aw32, B8, Bw35	3,4	+	+
B.F.	A11, Aw32, B13, B15	5,7	-	+
A.M.	A2, A3, B7, B18	1,2	+	+
J.S.	A3, A24, B7, B17	4,7	-	+
L.M.	A1, A2, B8, B15	NT	+	-
M.M.	A1, A3, B7, Bw52	NT	+	-
F.Y.	A3, A26, B5, B38	NT	+	+
B.G.	A1, A29, B14, B17	NT	+	+
D.C.	A2, A3, B7, B-	2,3	-	-
L.A.	A1, A2, B8, Bw40	1,3	-	-

PHA-activated lymphocytes from a selected panel of HLA-A1, HLA-A3 positive individuals. Those listed were selected from our regular HLA-screened panel and have been typed several times for study in two international workshops. NT, not tested.

anti-Ia (p29, 34) serum and subsequently testing K and SF48 alloantisera in a microcytotoxicity assay¹² (Table 1). This turkey antiserum, which does not bind mammalian complement, completely eliminated the reactivity with eight specific anti-DRw allosera (data not shown) and a rabbit anti-p29,34 serum C91 (Table 1). However, the reactivity of K and SF48 alloantisera was not affected by prior blocking with the turkey anti-Ia serum. These results indicated that the reactivity of K and SF48 alloantisera is not directed against DRw Ia-like determinants on the activated lymphocytes.

Taken together, these experiments indicated that PHA- and Con A-activated lymphocytes contain antigenic determinants that are associated with β_2 -microglobulin but are different from the classical HLA-A, -B and -C antigens. Serological studies on a random panel showed that these determinants are genetically related to the -A1 and -A3 alleles of the HLA-A locus, but as shown in Table 3, some false negative and false positive assignments can be made, indicating a serological independence of the K and SF48 specificities from the HLA-A1 and -A3 determinants. Genetic analyses of four families with informative segregation for both SF48 and K in microcytotoxicity assays indicated linkage of these determinants with HLA. One of these families (Fig. 1) demonstrated linkage between the locus coding for HLA-A and that coding for the antigenic determinant(s) recognised by SF48 alloantiserum.

Table 4 Cross-absorption with PHA-activated lymphocytes

Antisera	Absorbed with:		
	A3 ⁺ SF48 ⁺	A3 ⁺ SF48 ⁻	A3 ⁻ SF48 ⁻
Anti-HLA-A3	1	1	8
SF48	1	6	6
Testing cell = A3 ⁺ SF48 ⁺			

All absorptions and cytotoxicity scoring are as in Table 1. HLA phenotype of activated lymphocytes from the A3⁺, SF48⁺ individual D.W. is A1, A3, B8, B18, Cw2, from the A3⁺, SF48⁻ individual J.S. is A3, A24, B7, B17, and from the A3⁻, SF48⁻ individual D.F. is A11, Aw32, B14.1, B15.1, Cw3, Bw6.

We conclude that alloantisera K and SF48 detected antigenic determinants of a new polymorphic locus closely linked to the HLA-A locus on chromosome 6 in man. We have tentatively named it HT (human T locus) and have described two possible alleles, HT-1 and HT-3, based on the high degree of association and/or linkage disequilibrium with HLA-A1 and HLA-A3.

We thank Dr Richard J. Mahoney for reviewing the manuscript, Ms Rita Zimmermann and Kenneth LeClair for technical assistance and Ms K. Broadstreet and Mr B. Baillie for preparing the manuscript. This research was supported by NIH grants (AI 15066, CA 06516, CA 20531 and CA 19589 and by the Northeast Regional Red Cross Blood Program, Boston, Massachusetts.

Received 15 November 1979; accepted 10 January 1980.

- Klein, J. *et al.* *Immunogenetics* **6**, 489-512 (1978).
- Hammerling, G. J., Hammerling, U. & Flaherty, L. J. *exp. Med.* **150**, 108-116 (1979).
- Michaelson, J., Flaherty, L., Vitetta, E. & Poulik, M. J. *exp. Med.* **145**, 1066-1070 (1977).
- Tada, N., Tanigaki, N. & Pressman, D. J. *Immun.* **120**, 513-519 (1978).
- Ziegler, A. & Milstein, C. *Nature* **279**, 243-244 (1979).
- Terhorst, C., Parham, P., Mann, D. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 910-914 (1976).
- Ostberg, L., Rask, L., Nilsson, K. & Peterson, P. *Eur. J. Immun.* **5**, 462-468 (1975).
- Barnstable, C. J. *et al.* *Cell* **14**, 9-20 (1978).
- Evans, R. L. *et al.* *J. exp. Med.* **148**, 1440-1445 (1978).
- Fu, S. M. *et al.* *J. exp. Med.* **148**, 1423-1428 (1978).
- DeWolf, W. C., Schlossman, S. F. & Yunis, E. J. *J. Immun.* **122**, 1780-1784 (1979).
- Bodmer, W. F., Tripp, M. & Bodmer, J. in *Histocompatibility Testing* (eds Curtoni, E. S., Mattiuz, P. L. & Tosi, R. M.) 341-350 (Munksgaard, Copenhagen, 1967).
- Hoffman, T. & Kunkel, H. G. in *In Vitro Methods in Cell Mediated and Tumor Immunity* (eds Bloom, B. R. & David, J. R.) 71-81 (Academic, New York, 1976).
- Bernoco, D. *et al.* *Tissue Antigens* **8**, 256-260 (1976).
- Cejka, J. & Kithier, K. *Archs Biochem. Biophys.* **153**, 854-857 (1972).
- Crumpton, M. J. & Snary, D. *Contemp. Topics Mol. B.* **3**, 27-56 (1974).
- Robb, R., Terhorst, C. & Strominger, J. J. *biol. Chem.* **253**, 5319-5322 (1978).
- Springer, T. A. *et al.* *Nature* **268**, 213-218 (1977).

T-cell-derived helper factor allows *in vivo* induction of cytotoxic T cells in *nu/nu* mice

Hermann Wagner, Conny Hardt, Klaus Heeg, Martin Rölinghoff & Klaus Pfizenmaier

Institut für Medizinische Mikrobiologie der Johannes Gutenberg-Universität Mainz, Hochhaus am Augustusplatz, D-6500 Mainz, FRG

T-cell immunocompetence and diversity are thought to be generated in the thymus^{1,2}. This view is based on the findings that (1) T-cell ontogeny is thymus dependent^{3,4}, (2) the major histocompatibility restrictions of T-cell interactions are phenotypically related to the H-2 type of the thymus⁵⁻⁹, and (3) the phenotypic manifestation of H-2-linked immune responsiveness parallels the restriction elements selected in thymus¹⁰⁻¹². However, it is unclear whether pre-thymic cells programmed to develop into T cells do already express a receptor diversity, also whether pre-thymic cells have the potential to react against self-antigens, and whether the mechanism of self-tolerance is initiated in the thymus by either elimination or suppression of self-reactive clones. If it were possible to confer on pre-thymic cells antigen-specific effector functions, the impact of the thymus on the generation of T-cell diversity and function could be analysed in more detail. In mice, the nude mutation lacks a functioning thymus^{13,14}; *nu/nu* mice possess a thymic rudiment which is epithelial in the embryo¹⁵ and a fibrous, cystic remnant in adult^{15,16}; this remnant is not populated by lymphoid cells¹⁵⁻¹⁷. At present, the absence of immunocompetent T cells in *nu/nu* mice is explained by a lack of thymic differentiation and maturation of pre-thymic cells (reviewed in ref. 13). Here we report that injection of allogeneic stimulator cells plus a Lyt 1 T-cell-derived helper factor^{18,19}, termed interleukin 2 (for the system of nomenclature, see ref. 20) allows lymphocytes of *nu/nu* mice to differentiate *in vivo* into alloreactive cytotoxic T lymphocytes (CTLs).

The rationale of our experiments is based on the following observations. Murine thymocytes generate *in vitro* little, if any, cytolytic activity in response to otherwise immunogenic stimuli. However, in the presence of interleukin 2, the cytolytic response generated to alloantigens is almost equal to that normally existing in peripheral T cells²¹. More recently, we noted that even peanut-agglutinin-positive thymocytes, assumed to be cortical cells and thus immunoincompetent²², contained high numbers of CTL precursors which develop into CTLs in the presence of interleukin 2 (ref. 23). This suggested that the limiting factor in antigen-induced cytotoxic responses of immature thymocytes is not a lack of antigen-reactive CTL precursor cells, but a relative lack of cells producing interleukin

2. As cortical thymocytes had been assumed to be immunoincompetent, but exhibited immunocompetence in the presence of interleukin 2 (ref. 23), it was of interest to test whether in the presence of alloantigen interleukin 2 could also confer T-cell competence to lymphocytes in thymus-deficient *nu/nu* mice. The second observation relevant in this context was that sera of thymus-bearing normal mice inhibited the *in vitro* activity of the Lyt 1 T-cell-derived interleukin 2. In contrast, sera from athymic mice supported its functional activity (C.H., unpublished data). It was thus reasonable to assume that on injection of interleukin 2 into athymic mice, its functional activity would be maintained. Consequently, we tested whether *in vivo* T-cell precursors of *nu/nu* mice would respond to alloantigens, provided semi-purified¹⁸ interleukin 2 was simultaneously injected.

Of the various experimental approaches tested so far, reproducible results were obtained using the following protocols. Protocol A: On two consecutive days, 2×10^7 irradiated (2,000 rad) C57BL/6 mouse-derived splenic stimulator cells (H-2^b) were injected into the hind footpad of BALB/c *nu/nu* mice. Starting at the same time, the mice received, subcutaneously into the hind footpad, 100 μ l of helper factor twice a day for 3 d. Protocol B: BALB/c *nu/nu* mice received at the same time intervals the same dose of stimulator cells and helper factor intravenously. At day 5, the mice were killed and their lymph node cells and splenic cells tested for cytotoxic activity towards H-2^b, H-2^k and H-2^d target cells. As can be seen from Table 1, both protocols resulted in the *in vivo* induction of antigen-specific cytolytic effector cells within the spleen cells of *nu/nu* mice.

We subsequently defined a more convenient protocol for generating alloreactive cytolytic effector T cells from *nu/nu*-derived precursor T cells. Because the Lyt 1 T-cell-derived factor is generated in the course of a mixed lymphocyte culture, we reasoned that the factor ought also to be produced in the course of a graft-versus-host reaction (GvH) *in vivo*. Moreover, Piguet and Vasalli²⁴ recently reported that *nu/nu* mice are apparently able to reject up to 6×10^7 grafted allogeneic cells, provided the grafted cells contained T lymphocytes. This apparent paradoxical finding was not pursued further by the authors. One explanation for this is that during a T-cell-dependent GvH reaction induced in nude mice, helper factor may be produced *in vivo*, which in turn would allow *nu/nu* T-cell precursors to respond to the immunogenic stimulus represented by the grafted allogeneic cells.

In agreement with the results of Piguet and Vasalli²⁴, we noted that 8–9 d after injection of 6×10^7 allogeneic spleen cells the recipient *nu/nu* mice had rejected the grafted cells, that is, no grafted cells were detectable by serological techniques in the spleen of the killed recipient *nu/nu* mice (data not given). More interestingly, 8 d after grafting, the recipient *nu/nu* spleen cells exhibited low but significant cytolytic activity with specificity for target cells expressing the same alloantigen as the grafted cells (Table 2). The magnitude of cytolytic activity could be enhanced by restimulation of the *nu/nu* spleen cells *in vitro* in the presence

Table 1 *In vivo* induction of antigen-specific cytolytic effector cells in thymus-deficient nude mice

Protocol used	Cells tested	EL4 (H-2 ^b)			% Specific lysis of target cells			P815 (H-2 ^d)		
		500:1	120:1	30:1*	500:1	120:1	30:1	500:1	120:1	30:1
A	LN	12	9	2	1	0	1	–3	1	0
	spleen	31	10	3	0	–1	–1	8	6	3
B	LN	7	3	3	1	0	–1	–3	1	N.D.
	spleen	22	10	2	–1	3	2	2	–2	1

Protocol A: BALB/c nude mice (Bomholtgaard) were injected on two consecutive days, each with 2×10^7 C57BL/6-derived, irradiated (2,000 rad) splenic lymphocytes into the hind footpad. Simultaneously, the mice were injected at the same site, twice a day for 3 d, with 100 μ l semi-purified T-helper factor¹⁸. Protocol B: at identical time intervals, BALB/c nude mice received intravenously the same dose of stimulator cells and T-helper factor as in A. After 5 d, the mice were killed and single cell suspensions were prepared from either lymph node cells (LN) or spleen cells¹⁸. Lymphocytes were assayed for cytolytic activity in a 4-h standard ⁵¹Cr-release assay towards EL4 (H-2^b), LS (H-2^k) and P815 (H-2^d) target cells essentially as described elsewhere¹⁸. For each population tested, a dose-response curve was established. Per cent specific ⁵¹Cr-release was calculated according to the formula described previously¹⁸. Background lysis of the target cells was less than 23%.

* Ratio of effector cells to target cells used in the 4-h ⁵¹Cr-release assay.

Table 2 GvH reaction-induced sensitisation of alloreactive CTL in nude mice

Effector cells from	<i>In vivo</i> restimulation (5 d)	Treatment of effector cells	P815 (H-2 ^d)			% Specific lysis of targets EL4 (H-2 ^b)			LS (H-2 ^k)		
			300:1	60:1	12:1*	300:1	60:1	12:1	300:1	60:1	12:1
BALB/c (nudes) grafted with C57BL/6 spleen cells (H-2 ^b) for 8 d	ND	ND	4	2	—	15	13	6	1	-1	—
	H-2 ^b stimulator cells plus T-helper factor	NMS + C'	9	2	0	59	48	23	-1	0	—
		anti-Thy 1 + C'		ND		3	0	0		ND	
BALB/c (nudes) grafted with CBA spleen cells (H-2 ^k) for 8 d	ND	ND	4	0	2	3	2	-1	16	9	3
	H-2 ^k stimulator cells plus T-helper factor	NMS + C'	0	0	0	4	3	0	25	19	6
		anti-Thy 1 + C'		ND			ND		1	0	1

Thymus-deficient BALB/c mice received intravenously either 6×10^7 C57BL/6 mouse (H-2^b) or CBA mouse (H-2^k) splenic lymphocytes. After 8 d, the mice were killed and single cell suspensions were prepared from the spleen cells. According to the results obtained in complement-dependent cytotoxicity assay using anti-H-2^k, anti-H-2^d, anti-H-2^b antisera⁸, more than 95% of the spleen cells were BALB/c-mouse derived. One part of the spleen cells was assayed for cytotoxicity against ⁵¹Cr-labelled P815 (H-2^d), EL4 (H-2^b) and LS (H-2^k) target cells in a 4-h ⁵¹Cr assay. The remaining spleen cells (4×10^6) were cultured *in vitro* for an additional 5 d in the presence of irradiated stimulator cells (1×10^6) plus 50 μ l semi-purified T-helper factors¹⁸ in conditions described previously^{18,23}. Thereafter, the cells were treated with normal AKR serum plus complement or with monoclonal anti-Thy 1.2 antibody (derived from clone F7D5; Olac) plus C. The surviving cells were assayed for cytotoxicity against the target cells listed. Background lysis of the target cells during the 4 h ⁵¹Cr-assay did not exceed 20%.

* Ratio of effector to target cells. ND, not done.

of T-helper factor (Table 2). As the cytolytic activity was abrogated after treatment of the cells with anti-Thy 1 antisera plus complement (Table 2), we concluded that the antigen-specific cytolytic activity generated *in vivo* was carried out by *nu/nu* mouse-derived T cells.

Despite the presence of some Thy 1-positive cells²⁵ and possibly ectopic thymus tissue²⁶, normal *nu/nu* mice exhibit no T-cell functions^{27,28}. Although an increased *in vitro* development of Thy 1 antigen on the surface of nude spleen cells has been described following treatment with thymus extracts^{29,30}, culturing on thymus reticuloendothelial cell monolayers³¹ or exposing the cells to neuraminidase³² or poly(AU) (ref. 33), appearance of the Thy 1 surface antigen should only be equated with T-cell differentiation if T-cell function is demonstrated. This has been achieved recently by Gillis *et al.*³⁴. Accordingly, lymphocytes from *nu/nu* mice were able to differentiate *in vitro* into alloreactive CTLs, provided the cells were cultured in the presence of stimulator cells plus T-cell growth factor (TCGF). Obviously, these *in vitro* results almost parallel our *in vivo* findings described here.

Our results are relevant to two issues. First, a well-characterised lymphokine^{18,19} is now shown to be active *in vivo* in restoring T-cell immune functions in *nu/nu* mice. Second, the data imply that *nu/nu* mice contain lymphoid cells which can be activated *in vivo* by an allogeneic stimulus and interleukin 2 to generate cytotoxic T cells.

Recent work has shown that interleukin 2 is released from Lyt 1 T cells in response to allogeneic and mitogenic stimuli¹⁸. *In vitro*, interleukin 2 replaces the Lyt 1 T-helper cell requirement of thymic Lyt 123 CTL precursors, thus allowing the latter to differentiate in response to antigenic stimuli into alloreactive of H-2-restricted Lyt 23 CTLs (ref. 23 and C.H. *et al.*, in preparation). Systemic *in vivo* application of interleukin 2 plus alloantigen is shown here to trigger in otherwise T-cell immunoincompetent *nu/nu* mice the induction of alloreactive CTLs. Therefore, *nu/nu* mice must possess interleukin 2-sensitive T-precursor cells endowed with the receptor repertoire required for alloreactivity.

One could argue that the thymic rudiment of *nu/nu* mice induces T-cell lymphopoiesis with a functional activity below normal detection. This low activity might be boosted by the *in vivo* activity of interleukin 2 to the levels described here. If so, *in vivo* application of interleukin 2 might be useful as an immunopharmacological agent for amplifying weak T-cell reactivity *in vivo*. If, however, the thymus defect in *nu/nu* mice is absolute, the results imply that interleukin 2 bypasses the requirement of a thymus during the *in vivo* differentiation of pre-thymocytes into

alloreactive CTLs. According to this reasoning, pre-thymic cells already contain the receptor repertoire required for alloreactivity. We do not know whether the repertoire for alloreactivity is dependent on pre-thymic somatic events or whether it is inherited from the germ line, as suggested by Jerne². Because the repertoire for alloreactivity might already be expressed on pre-thymic cells, and because alloreactive and H-2-restricted CTLs may be identical^{35,36}, it is also possible that the repertoire for extrinsic foreign antigens is expressed in pre-thymic precursor T cells. If so, we envisage, at the level of precursor T cells, distinct clones with the potential to react to an extrinsic antigen in the context of both syngeneic and allogeneic H-2 structures.

We thank Sabine Barth, Monika Schmidt and Rosi Trabschitt for technical help. This work was supported by the SFB 107, and the Stiftung Volkswagenwerk.

Note added in proof: Using the very same protocol we recently succeeded in inducing in *nu/nu* mice T helper cells specific for heterologous sheep erythrocytes (Stötter, Rude and Wayne, submitted for publication).

Received 16 October 1979; accepted 14 January 1980.

1. Burnet, F. M. *Immunological Surveillance* (Pergamon, Sydney, 1970).
2. Jerne, N. K. *Eur. J. Immun.* **1**, 1-9 (1971).
3. Miller, J. F. A. P. & Osoba, D. *Physiol. Rev.* **47**, 437-478 (1967).
4. Davies, A. J. S., Lenchars, E., Wallis, V., Marchant, R. & Elliot, E. V. *Transplantation* **5**, 222-231 (1967).
5. Doherty, P. C., Blanden, R. V. & Zinkernagel, R. M. *Transplant. Rev.* **29**, 89-124 (1976).
6. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 822-896 (1978).
7. Bevan, M. J. *Nature* **269**, 417-418 (1978).
8. Pfizenmaier, K., Starzinski-Powitz, A., Rodt, H., Rölinghoff, M. & Wagner, H. *J. exp. Med.* **143**, 999-1004 (1976).
9. Gordon, R. D., Simpson, E. & Samelson, L. E. *J. exp. Med.* **142**, 1108-1120 (1975).
10. v. Boehmer, H., Haas, W. & Jerne, N. K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2439-2441 (1978).
11. Matsunaga, T. & Simpson, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6207-6209 (1978).
12. Zinkernagel, R. M., Althage, A., Cooper, S., Callahan, G. N. & Klein, J. *J. exp. Med.* **148**, 805-810 (1978).
13. Kindred, B. *Prog. Allergy* **26**, 137-238 (1979).
14. Pantelouris, E. M. *Nature* **217**, 370-371 (1968).
15. Jordan, R. K., Owen, J. J. T. & Raff, M. C. *Eur. J. Immun.* **7**, 736-743 (1977).
16. Owen, J. J. T., Jordan, R. F. & Raff, M. C. *Eur. J. Immun.* **5**, 653-656 (1975).
17. Cordier, A. C. *J. ultrastruct. Res.* **47**, 26-32 (1974).
18. Wagner, H. & Rölinghoff, M. *J. exp. Med.* **148**, 1523-1528 (1977).
19. Watson, J., Gillis, S., Marbrook, J., Mochizuki, D., & Smith, K. A. *J. exp. Med.* **150**, 849-861 (1979).
20. *2nd Int. Workshop*, Ermatingen (1979).
21. Wagner, H., Rölinghoff, M., Schwallier, R., Hardt, C., & Pfizenmaier, K. *Nature* **280**, 5721-5722 (1979).
22. Irlé, C., Piguet, P. F. & Vasalli, P. *J. exp. Med.* **148**, 32-45 (1978).
23. Wagner, H., Rölinghoff, M., Pfizenmaier, K., Hardt, C. & Johnscher, G. *J. Immun.* (in the press).
24. Piguet, P. F. & Vasalli, P. *J. Immun.* **120**, 79-83 (1978).
25. Loo, F., Amstutz, H., Hägg, L. B., Major, K. S. & Rochants, G. E. *Eur. J. Immun.* **6**, 663-667 (1976).
26. Hair, J. in *Proc. 1st Int. Workshop in Nude Mice*, 23 (Fischer, Stuttgart, 1974).
27. Feldman, M., Wagner, H., Basten, A. & Holmes, M. *Aust. J. exp. Biol. med. Sci.* **50**, 651-660 (1972).
28. Wagner, H. *J. Immun.* **109**, 630-635 (1972).
29. Komuro, K. & Boyse, E. A. *J. exp. Med.* **138**, 479-483 (1973).

30. Scheid, M. P., Golstein, G. & Boyse, E. A. *Science* **190**, 2111–2112 (1974).
 31. Sato, V. L., Waksal, S. D. & Herzenberg, L. A. *Cell. Immun.* **24**, 173–179 (1976).
 32. Milewicz, C., Miller, H. G. & Esselman, W. J. *J. Immun.* **117**, 1774–1779 (1976).
 33. Scheid, M. P. *et al.* *J. exp. Med.* **138**, 1027–1031 (1973).
 34. Gillis, St., Union, N. A., Baker, P. E. & Smith, K. A. *J. exp. Med.* **149**, 1460–1476 (1979).
 35. Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2094–2098 (1977).
 36. Finberg, R., Burakoff, St., Cantor, H. & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5145–5149 (1978).

Chemical characterisation of the Fab and Fc fragments from surface immunoglobulin

R. M. E. Parkhouse*, John Lifter† & Y. S. Choi‡

*Medical Research Council, Mill Hill, London NW7 1AA, UK

†Department of Internal Medicine, Yale University, New Haven, Connecticut 06510

‡Sloan-Kettering Institute for Cancer Research, 145 Boston Post Road, Rye, New York 10580

Immunoglobulin (Ig) molecules of the M and D classes are present on the membranes of B lymphocytes (sIg), where they serve as antigen receptors^{1–6}. sIg is a biosynthetically stable membrane protein which requires either denaturing conditions or detergents to free it from other membrane constituents^{1,7–9}. Whether the tight association of sIg with the membrane is due to the chemical nature of the sIg molecule itself or to another constituent of the membrane, and the mechanism by which the recognition of antigen becomes a signal for cell division and differentiation, have become topics of current interest. Several mechanisms by which the membrane–sIg association could be maintained have been proposed^{10–13}: the sIg could have a unique peptide structure and/or amino acid sequence responsible for stabilising the association; the carbohydrate composition of sIg could impart physical and/or chemical characteristics which maintain the association; sIg could be closely associated with another protein, the proreceptor, which would be responsible for holding the sIg in the membrane. Here, we use charge-shift electrophoresis¹⁴ to examine the detergent-binding properties of mouse sIgM and sIgD and the tryptic Fab and Fc fragments derived from these molecules. In this technique, hydrophobic proteins which bind detergent are recognised by an altered electrophoretic behaviour in the presence of charged detergents. Proteins which fail to bind detergent do not show a 'shift' in electrophoretic migration. Our results confirm the observation that sIg is a hydrophobic protein which binds detergent^{15–17}, and demonstrate that the site of detergent binding is located in the C-terminal portion of the sIg heavy chain.

Suspensions of murine splenocytes were radioiodinated and the labelled surface proteins solubilised by detergent lysis^{18,19}. sIgM, sIgD and the Fab and Fc fragments were isolated by a combination of affinity chromatography and tryptic digestion (Fig. 1). The ¹²⁵I-labelled lysate was divided into two aliquots. One aliquot was passed over an anti-μ affinity column; complete sIgM molecules were subsequently eluted from this column. The unbound material contained ¹²⁵I-sIgD in addition to other labelled surface proteins. This eluate was digested with trypsin in conditions previously shown to give complete cleavage of sIgD into Fab and Fc fragments²⁰. The digest was applied to an anti-δ affinity column from which sIgD Fc fragments were eluted. The digest, depleted of Fc fragments, was passed over an anti-light-chain column from which the sIgD Fab fragments were subsequently obtained.

The second aliquot was subjected to a similar procedure except that the anti-δ affinity column preceded, and the anti-μ affinity column followed, the proteolytic digestion. The hydrophobicity of the sIg molecules and tryptic fragments was examined by charge-shift electrophoresis¹⁴ (Fig. 2) and gave the following results. Neither the myeloma proteins, run as standards, nor the sIg or sIg fragments moved from the origin in the presence of the non-ionic detergent Triton X-100. In the presence of the anionic detergent deoxycholate, anodal migra-

tion of the sIgM, sIgD, sIgM Fc and sIgD Fc fragments was observed.

These results demonstrated that the Fc fragments of sIg contain the hydrophobic regions responsible for detergent binding. Perhaps the simplest explanation for these observations is that sIgM and sIgD have a unique stretch of hydrophobic amino acids at, or near to, the C-terminus of the heavy chain. The recent experiments of Williams *et al.*²¹ support this hypothesis, although the earlier experiments of Walsh and Crumpton²² and McIlhinney *et al.*²³ lead to a contrasting, although unproven conclusion.

Another possibility is that the hydrophobicity was due to incomplete glycosylation of the sIg. Several arguments against a relationship between the extent of glycosylation and detergent binding can be made. First, removal of carbohydrate residues from secreted human IgM myeloma proteins with neuraminidase and β-galactosidase did not result in an appearance of detergent-binding properties (J.L., unpublished observation). Second, secreted IgG does not contain a sub-fraction which binds detergent¹⁴, although the carbohydrate composition of these molecules is known to be heterogeneous²⁴. Third, other integral membrane proteins, such as the murine and human histocompatibility antigens, are highly glycosylated, hydrophobic and capable of binding detergent. Finally, intracellular 7S IgM from mouse myeloma cells, although deficient in carbohydrate²⁵, does not bind detergent in the charge-shift electrophoresis assay (R.M.E.P., unpublished work).

It is possible that the hydrophobicity of sIg was actually due to a closely associated, noncovalently bound hydrophobic proreceptor undetected by the iodination procedures. This possibility is less likely in view of the observations that biosynthetically radiolabelled intracellular Ig from the murine B-cell leukaemia 70Z/3 (J.L. and C. J. Paige, manuscript in preparation) or from chicken bursal cells²⁶ showed the same electrophoretic behaviour as we report for murine sIgM and sIgD. This biosynthetically labelled Ig was free of peptides other than heavy and light chains when the cells were solubilised with both NP-40 and 2% sodium deoxycholate^{13,27}.

Although not conclusive, our results are consistent with the interpretation that the hydrophobic nature of sIg is sufficient to account for the stability of the plasma membrane–sIg asso-

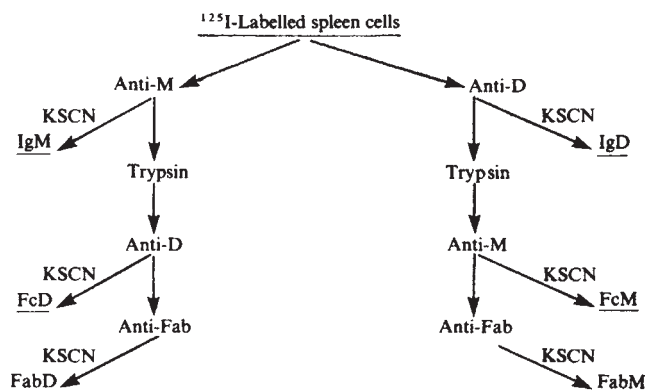


Fig. 1 Preparation of complete and tryptic cleaved sIg. The lysate was divided into two aliquots and subjected to the first stage of affinity chromatography as described in the text. The sIgD-containing lysate was digested with TPCK–trypsin (Sigma) at 0.1 mg ml⁻¹ for 30 min at room temperature. The sIgM-containing lysate was treated identically except that incubation was at 37 °C for 3 h. The digestion was stopped by the addition of 1 mg ml⁻¹ soybean trypsin inhibitor, 1 mg ml⁻¹ egg white trypsin inhibitor and 1 mM phenylmethylsulphonyl fluoride. The second and third stages of affinity chromatography followed. Material bound to the affinity resins was eluted with 4 M KSCN containing 0.5% Triton X-100 after washing the column with 0.5% Nonidet P-40 phosphate-buffered iodine, pH 7.4, and four column volumes of 1 M KSCN containing 0.5% Triton X-100. The samples were dialysed against 0.05 M glycine-NaOH, 0.1 M, NaCl and 0.5% Triton X-100, pH 9.0.

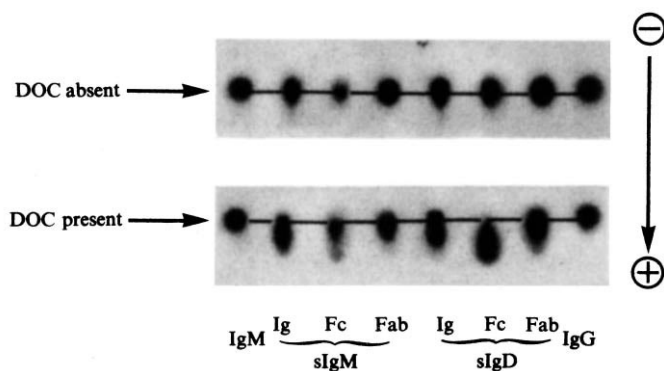


Fig. 2 Charge-shift electrophoresis of sIg and tryptic fragments. Rabbit antisera to mouse μ -chains, δ -chains and κ -chains²⁸ were precipitated at 1.6 M ammonium sulphate and the immunoglobulin fractions coupled to CH-Sepharose 4B (Pharmacia) using carbodiimide according to the manufacturer's instructions. The anti- μ and anti- δ sera were without specificity towards the Fab portions of the molecules. Spleen cells from male CBA mice were radioiodinated by the lactoperoxidase-catalysed procedure and lysed in 0.5% (w/v) Nonidet P-40^{18,19}. The nuclei and cellular debris were removed by centrifugation at 15,000g for 15 min, 0 °C, and the free ¹²⁵I-labelled serum IgM and IgG samples were purified mouse myeloma proteins MOPC 104E and Adj PC5, respectively. The migration of these samples with or without sodium deoxycholate was unaffected by prior chemical reduction with dithiothreitol or treatment with 4 M KCNS and dialysis against the electrophoresis buffer.

Inherited defect in a Na⁺, K⁺-co-transport system in erythrocytes from essential hypertensive patients

Ricardo P. Garay, Georges Dagher, Marie-Gabrielle Pernollet, Marie-Aude Devynck & Philippe Meyer

U7 INSERM/ERA 336 CNRS, Hôpital Necker, 161 rue de Sèvres, 75015 Paris, France

The Na⁺ and K⁺ electrochemical gradients across cell membranes are believed to be maintained by the action of a Na⁺, K⁺-pump¹⁻³. In human erythrocytes this pump exchanges internal Na⁺ for external K⁺ in approximately a 1.5 ratio^{4,5}. Thus, when Na⁺-loaded/K⁺-depleted erythrocytes are incubated in physiological conditions they tend to recover their original low Na⁺/high K⁺ content. Surprisingly, in erythrocytes from healthy donors the net Na⁺ extrusion/K⁺ influx ratio exceeds the 1.5 ratio predicted for Na⁺, K⁺-pump-mediated fluxes whereas it is similar to this value in erythrocytes from essential hypertensive patients and some of their descendants⁶. We now report that this difference is due to the presence of a Na⁺, K⁺-co-transport system in normal erythrocytes which extrudes both internal Na⁺ and K⁺ and is functionally deficient in erythrocytes of essential hypertensive patients and some of their descendants. No difference in passive Na⁺ permeability could be detected between normotensive and hypertensive subjects.

The different transmembrane pathways involved in net Na⁺ and K⁺ transport in Na⁺-loaded/K⁺-depleted erythrocytes were analysed in the following groups⁶: normotensive and secondary hypertensive subjects devoid of familial hypertension, benign and accelerated essential hypertensive patients and those offspring of hypertensive patients with an abnormal low ratio of Na⁺/K⁺ net fluxes (⊕ normotensives).

The presence of ouabain (0.1 mM) in the incubation medium greatly reduced net Na⁺/K⁺ movements in all subjects studied (Table 1). If we assume that the ouabain-sensitive fraction of the total fluxes corresponds to the exchange of internal Na⁺ for external K⁺ catalysed by the Na⁺, K⁺-pump (see ref. 7 for discussion), three conclusions can be drawn:

First, in erythrocytes of ⊕ normotensives and benign essential hypertensive patients the pump fluxes are 20–40% higher than in normal erythrocytes. A high (Na⁺ + K⁺)ATPase activity has been recently reported for benign essential hypertensives⁸. This increased pump activity could result either from a more rapid rate of cation translocation or from a higher density of pump unities. To test this latter hypothesis we have studied the binding of ³H-ouabain which is known to label the (Na⁺ + K⁺)ATPase specifically. Scatchard analysis of ³H-ouabain binding to erythrocyte membranes indicated the presence of only one class of binding sites, the number of which did not significantly differ between the five groups (Table 1). In contrast, the apparent dissociation constant (K_D) varied from 3.7×10^{-9} M to 1.3×10^{-8} M. A correlation was found between the apparent dissociation constant and the pump activation (Fig. 1). It is interesting that a regulatory membrane protein changing the apparent affinity for the ouabain inhibition of the (Na⁺ + K⁺)ATPase has been described recently⁹. This phenomenon certainly merits further investigation.

Second, in all subjects the stoichiometry of the Na⁺, K⁺-pump is very close to that reported previously^{4,5}, that is, ouabain-sensitive Na⁺ efflux/K⁺ influx = -1.5; the minus sign indicates that the vectorial transports of Na⁺ and K⁺ have different signs.

Third, in the presence of 0.1 mM ouabain a net Na⁺ extrusion in an uphill direction against the electrochemical Na⁺ gradient was observed in erythrocytes of normotensive controls (and

ciation. The contribution of the Fc-localised hydrophobic region of sIg to the mechanism of signal transduction, whether the sIg is a transmembrane protein, and the structural differences between sIg and serum immunoglobulin of the same class, remain to be determined.

This work was supported by NCI grants CA-25758 (J.L.) and CA-18613 (Y.S.C.). The technical assistance of Mrs Gillian Clark is appreciated.

Received 16 November 1979; accepted 28 January 1980.

- Vitetta, E. S., Baur, S. & Uhr, J. W. *J. exp. med.* **134**, 242–264 (1971).
- van Boxel, J. A., Paul, W. E., Terry, W. D. & Green, I. J. *Immun.* **109**, 648–651 (1972).
- Rowe, D. J., Hug, K., Faulk, W. P., McCormick, J. N. & Gerber, H. *Nature new Biol.* **242**, 155–156 (1973).
- Fu, S. M., Winchester, R. J. & Kunkel, H. G. *J. exp. Med.* **139**, 451–456 (1974).
- Melcher, U. *et al. J. exp. med.* **140**, 1427–1431 (1974).
- Abney, E. R. & Parkhouse, R. M. E. *Nature* **252**, 600–602 (1974).
- Singer, S. J. & Nicolson, G. L. *Science* **175**, 720–731 (1972).
- Choi, Y. S. *Biochemistry* **15**, 1037–1042 (1976).
- Lifter, J., Kincade, P. W. & Choi, Y. S. *J. exp. Med.* **117**, 2220–2225 (1976).
- Melchers, F. *Biochemistry* **11**, 2204–2208 (1973).
- Ramasamy, R., Munro, A. & Milstein, C. *Nature* **249**, 573–574 (1974).
- Ramasamy, R. *Immunochimistry* **13**, 705–708 (1976).
- Lifter, J. & Choi, Y. S. in *Avian Immunology* (ed. Benedict, A. A.) 99–107 (Plenum, New York, 1977).
- Helenius, A. & Simons, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 529–532 (1977).
- Melcher, U., Eidels, L. & Uhr, J. W. *Nature* **258**, 434–435 (1975).
- Melcher, U. & Uhr, J. W. *Biochemistry* **16**, 145–152 (1977).
- Melcher, U. *Immunochimistry* **15**, 483–489 (1978).
- Phillips, D. R. & Morrison, M. *Biochem. biophys. Res. Commun.* **40**, 284–289 (1970).
- Abney, E. R., Evans, W. H. & Parkhouse, R. M. E. *Biochem. J.* **159**, 293–299 (1976).
- Bourgeois, A., Abney, E. R. & Parkhouse, R. M. E. *Eur. J. Immun.* **7**, 210–213 (1978).
- Williams, P. B., Kubo, R. T. & Grey, H. W. *J. Immun.* **121**, 2435–2439 (1978).
- Walsh, F. S. & Crumpton, M. J. *Nature* **269**, 307–311 (1977).
- McIlhinney, R. A. J., Richardson, N. E. & Feinstein, A. *Nature* **272**, 555–557 (1978).
- Niedermeier, W., Kirkland, T. & Acton, R. T. *Biochim. biophys. Acta* **237**, 442–445 (1971).
- Parkhouse, R. M. E. & Melchers, F. *Biochem. J.* **125**, 235–240 (1971).
- Lifter, J. & Choi, Y. S. *Fedn Proc.* **37**, 1838 (1978).
- Singer, B. A. & Williamson, A. R. *Eur. J. Immun.* (in the press).
- Abney, E. R., Hunter, I. R. & Parkhouse, R. M. E. *Nature* **259**, 404–406 (1976).
- Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82**, 314–316 (1977).

Table 1 Characteristics of net Na⁺ and K⁺ transport in erythrocytes from normotensive and hypertensive patients

Flux component	Transport system	Net flux*	Absence of familial hypertension		Familial history of hypertension		
			Normotensive controls	Secondary hypertensives	⊕ Normotensives	Essential hypertensives	
			(n = 7)	(n = 4)	(n = 6)	(n = 8)	(n = 6)
Total	All systems	Na ⁺	-3.70 ± 0.24	-3.91 ± 0.68	-3.95 ± 0.87	-3.92 ± 0.81	-2.74 ± 0.39§
		K ⁺	+1.10 ± 0.35	+1.47 ± 0.22	+2.32 ± 0.59§	+2.16 ± 0.45§	+1.55 ± 0.28
		Na ⁺ /K ⁺	-3.59 ± 0.89	-2.67 ± 0.24	-1.72 ± 0.10§	-1.82 ± 0.27§	-1.78 ± 0.09§
Ouabain-sensitive	Na ⁺ , K ⁺ -pump	Na ⁺	-3.20 ± 0.23	-3.35 ± 0.48	-4.64 ± 0.51§	-4.52 ± 0.67§	-3.02 ± 0.48
		K ⁺	+2.08 ± 0.29	+2.13 ± 0.40	+2.87 ± 0.42‡	+2.79 ± 0.43‡	+2.12 ± 0.32
		Na ⁺ /K ⁺	-1.59 ± 0.19	-1.58 ± 0.25	-1.63 ± 0.15	-1.63 ± 0.19	-1.38 ± 0.18
Ouabain-resistant	Non-pumped fluxes	Oub N _s	(n = 9) 542 ± 183	(n = 2) 540 ±	(n = 8) 535 ± 185	(n = 7) 637 ± 63	(n = 3) 486 ± 119
		Na ⁺	-0.51 ± 0.17	-0.55 ± 0.49	+0.69 ± 0.40	+0.61 ± 0.30	+0.27 ± 0.21
		K ⁺	-0.98 ± 0.15	-0.66 ± 0.20	-0.55 ± 0.26	-0.63 ± 0.24	-0.57 ± 0.27
Ouabain-resistant furosemide-sensitive	Na ⁺ , K ⁺ -co-transport	Na ⁺	(n = 10) -0.52 ± 0.11		(n = 3) -0.10 ± 0.02‡	(n = 10) -0.14 ± 0.18§	
		K ⁺	-0.76 ± 0.27		-0.13 ± 0.08‡	-0.19 ± 0.16§	
		Na ⁺ /K ⁺	+0.73 ± 0.21				
Ouabain- and furosemide-resistant	Passive permeability†	Na ⁺	(n = 7) +0.25 ± 0.05			(n = 8) +0.26 ± 0.06	
		K ⁺	(n = 10) -0.31 ± 0.17			(n = 10) -0.31 ± 0.12	

The effect of ouabain on the net Na⁺ and K⁺ fluxes was studied by the method previously described⁶ with the following modifications: (1) the PCMBs concentration of the Na⁺-loading solution was reduced to 0.02 mM, (2) 2 mM adenine, 3 mM inosine and 2.9 mM Na⁺ phosphate buffer (pH 7.2 at 37° C) were added to the recovering solution. After recovery, the internal cation content was (mmol per l cells) 55–75 Na⁺ and 30–50 K⁺ and the increase in cell volume was less than 2%, (3) ouabain-sensitive fluxes were calculated by the difference between net Na⁺ and K⁺ fluxes in a Na⁺, K⁺-Ringer solution with and without ouabain 0.1 mM. At hours 0, 1, 2 and 3, two tubes were transferred to 0° C and the cells were washed three times with 150 mM choline Cl, 2.5 mM Tris-phosphate (pH 7.2 at 4° C) and 0.02 mM ouabain at 4° C. Na⁺, K⁺ and haemoglobin were measured in the same sample. To study the effect of furosemide on the ouabain-resistant net Na⁺ and K⁺ fluxes some additional modifications were introduced: (1) the Na⁺-loading solution was (mM) 250 choline Cl, 40 NaCl, 10 KCl, 2.5 Na phosphate (pH 7.2 at 4° C), 1 MgCl₂ and 0.02 PCMBs, (2) the recovering solution was K⁺ free. After recovery the reduction in cell volume was less than 3% and the internal cation content was 20–30 mmol per l cells of both Na⁺ and K⁺. Furosemide-sensitive fluxes were calculated by the difference between net Na⁺ and K⁺ fluxes in a medium containing (mM) 90 choline Cl, 60 NaCl, 2.5 Na phosphate (pH 7.2 at 37° C), 1 MgCl₂, 10 glucose and 0.1 ouabain with and without 1 mM furosemide. At each incubation time three tubes were removed from the incubation bath. Ouabain + furosemide-resistant net Na⁺ fluxes were measured in fresh cells. After removing the plasma and buffy coat, the cells were washed twice with 150 mM choline Cl. They were then incubated at 37° C in three different Na⁺-media containing (mM) 1 MgCl₂, 2.5 Tris-phosphate (pH 7.2 at 37° C), 10 glucose, 0.1 ouabain, 1 furosemide and 0, 75 or 150 NaCl (the isotonicity was maintained with choline Cl). At hours 0, 1 and 2, two tubes were transferred to 0° C and the cells were processed as before. The rate constant of Na⁺ permeability was obtained from the slope relating the ouabain- and furosemide-resistant net Na⁺ fluxes against the external Na⁺ concentration. Ouabain- and furosemide-resistant net K⁺ fluxes were measured in the same PCMBs-treated cells used for measuring furosemide-sensitive fluxes. Ouabain-binding experiments were carried out with washed erythrocytes incubated in (mM) 1 HEPES (pH 7.4), 150 NaCl, 3 ATP and 3 MgCl₂ for 3 h at 37° C with 5 × 10⁻¹⁰ M to 10⁻⁷ M ³H-ouabain (13 Ci mmol⁻¹, NEN). Results were corrected by subtracting the nonspecific binding measured in the additional presence of 10⁻⁶ M unlabelled ouabain. The values given represent mean ± s.d. and n denotes the number of subjects in each group. The statistical significance of the data was studied using the non-parametric test of Mann-Whitney.

* Net flux is given in mmol per l cells per h. A minus sign indicates a net cation extrusion, a plus sign a net cation influx. Na⁺/K⁺ values indicate the stoichiometry of Na⁺/K⁺ net fluxes, and Oub N_s the number of ouabain binding sites per cell.

† Rate constant of cation permeability in d⁻¹.

‡ P < 0.01; § < 0.001.

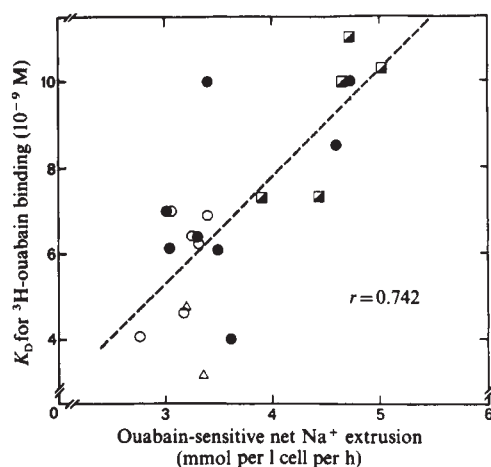


Fig. 1 Relationship between the Na⁺, K⁺-pump activation and the apparent dissociation constant of ouabain binding. ○, normotensive controls; ●, essential hypertensives; △, secondary hypertensives; ■, normotensives born of hypertensive parents.

secondary hypertensives). This confirms previous observations suggesting that the ouabain-resistant fluxes do not simply represent passive permeabilities¹⁰ and that the Na⁺, K⁺-pump is not the only mechanism ensuring a net cell Na⁺ extrusion^{11,12}. It is interesting that no ouabain-resistant net Na⁺ extrusion was observed in erythrocytes from ⊕ normotensives and essential hypertensive patients (Table 1).

In preliminary experiments we found that the pump-independent net Na⁺ extrusion of normal erythrocytes could be inhibited by raising the external K⁺ concentration, by adding 1 mM furosemide or by ATP depletion, thus showing properties similar to the Na⁺, K⁺-co-transport system described in rat¹³, avian¹⁴ and human erythrocytes^{15,16}. In certain experimental conditions, this system may extrude internal Na⁺ in an uphill direction against the electrochemical Na⁺ gradient, using the energy supplied by the downhill K⁺ extrusion across the electrochemical K⁺ gradient. This system was therefore investigated in experiments in which conditions were modified to measure satisfactorily the small ouabain-resistant net Na⁺ and K⁺ fluxes (see Table 1).

In the presence of 0.1 mM ouabain, the addition of 1 mM furosemide to normal erythrocytes loaded with Na⁺ and choline effectively inhibited a net Na⁺ and K⁺ extrusion coupled in a ratio slightly less than +1 (the plus sign indicates that vectorial transport of both ions has the same sign) (Table 1). In contrast,

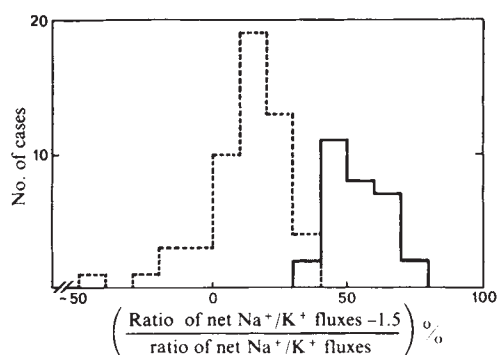


Fig. 2 Histograms of the ratio of total Na^+/K^+ net fluxes represented as the percentage deviation from the Na^+ , K^+ -pump stoichiometry. Continuous line represents normotensive controls (30 subjects) and dotted line essential hypertensive patients (54 subjects).

the net Na^+ and K^+ extrusion catalysed by this Na^+ , K^+ -co-transport system is markedly reduced in erythrocytes of essential hypertensive patients and $\oplus$ normotensives (Table 1). Although the Na^+ , K^+ -co-transport fluxes are smaller than the pump fluxes, the fact that they have a positive stoichiometry as opposed to the negative stoichiometry of the Na^+ , K^+ -pump (and of the passive net Na^+ and K^+ fluxes) introduces a marked difference between the stoichiometry of total net fluxes of essential hypertensives and normotensive controls (see Fig. 2).

Net Na^+ and K^+ fluxes in the presence of both ouabain (0.1 mM) and furosemide (1 mM) have the kinetic properties of passive 'leak' fluxes. In fresh red blood cells, the ouabain- and furosemide-resistant net Na^+ influx (1) increases linearly with the external Na^+ concentration, (2) is not modified by external K^+ and (3) becomes zero when the external Na^+ equals the internal Na^+ concentration (in mmol l^{-1} cell). Moreover, these properties do not seem to be modified by incubation with cold 0.02 mM 2,5-*p*-chloromercuribenzenesulphonate (PCMBS). As indicated in Table 1, no difference in the passive Na^+ permeability could be detected between erythrocytes of hypertensive patients and those of normotensive controls. The same is true for the passive K^+ permeability which was only measured in PCMBS-treated cells. Thus, the increase in the ouabain-resistant unidirectional Na^+ influx¹⁷ and Na^+ efflux¹⁸ previously reported in erythrocytes from essential hypertensives, rather than being due to an increased passive Na^+ permeability, might reflect a disorder in a specific transport system such as the ouabain-resistant one-to-one Na^+/Na^+ exchange which is activated in erythrocytes from essential hypertensive patients¹⁹. If this were the case, the increased unidirectional Na^+ fluxes would not be related to our observations because the Na^+/Na^+ exchange mechanism does not result in net Na^+ fluxes.

The results presented here indicate that the net extrusion of an erythrocyte Na^+ load is accomplished by two different mechanisms, the Na^+ , K^+ -pump and Na^+ , K^+ -co-transport, which operate against the passive net Na^+ fluxes. Furthermore, they suggest that an inherited defect in the Na^+ , K^+ -co-transport system may be genetically associated with essential hypertension. In avian red cells this transport system participates in the regulation of cell volume and is under hormonal control. These observations have led us to formulate the hypothesis that essential hypertension may result from an inefficient hormonal regulation of the extrusion of a cell Na^+ load after an excess Na^+ intake due to a functional disorder of the Na^+ , K^+ -co-transport system. Such a process in excitable cells of high surface/volume ratio, such as smooth muscle cells or catecholaminergic neurones, may lead to a temporary or permanent increase in intracellular Na^+ , producing critical changes capable of raising blood pressure^{20,21}. The high Na^+ , K^+ -pump activity seen in $\oplus$ normotensives and benign essential hypertensives may therefore represent a compensatory mechanism for extruding a cell Na^+ load and thus preventing severe hypertension in subjects with this genetic abnormality.

We thank Drs L. Beaugé, I. Glynn, V. Lew, N. Adragna, M. Cannessa, D. Tosteson and M. de Mendonça for interest and discussion, and Corinne Arnould and Alain Cardot for technical assistance. This study was supported by the CNRS and the INSERM.

Received 7 July 1979; accepted 17 January 1980.

- Glynn, I. & Karlish, S. A. *Rev. Physiol.* **37**, 13–55 (1975).
- Sarkadi, B. & Tosteson, D. in *Membrane Transport in Biology* Vol. 2 (eds Giebisch, G., Tosteson, D. & Ussing, H.) 117–160 (Springer, Heidelberg, 1979).
- Garrahan, P. & Garay, R. *Curr. Topics Membrane Transport* **8**, 29–97 (1976).
- Post, R. & Jolly, P. *Biochim. biophys. Acta* **25**, 118–128 (1957).
- Whittam, R. & Ager, N. *Biochem. J.* **97**, 214–227 (1965).
- Garay, R. P. & Meyer, P. *Lancet* **i**, 349–353 (1979).
- Lew, V. & Beaugé, L. in *Membrane Transport in Biology* (eds Giebisch, G., Tosteson, D. & Ussing, H.) 81–115 (Springer, Heidelberg, 1979).
- Wambach, G., Helber, A., Bönner, G. & Hummerich, W. *Klin. Wochenschr.* **57**, 169–172 (1979).
- Zachowski, A., Lelièvre, L., Aubry, J., Charlemagne, D. & Paraf, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 633–637 (1977).
- Glynn, I. J. *J. Physiol., Lond.* **136**, 148–173 (1957).
- Hoffman, J. & Kregenow, F. *Ann. N.Y. Acad. Sci.* **137**, 566–576 (1966).
- Sachs, J. J. *gen. Physiol.* **57**, 259–282 (1971).
- Beaugé, L. & Ortiz, O. *J. Membrane Biol.* **13**, 165–184 (1973).
- McManus, T. & Schmidt, W. in *Membrane Transport Processes* Vol. 1 (ed. Hoffman, J.) 79–106 (Raven, 1978).
- Beaugé, L. & Adragna, N. *J. gen. Physiol.* **57**, 576–592 (1971).
- Wiley, J. & Cooper, R. *J. clin. Invest.* **53**, 745–755 (1974).
- Wessels, F. & Losse, M. *Z. Kreislaufforsch.* **56**, 374–380 (1967).
- Postnov, Y., Orlov, S., Shevchenko, A. & Adler, A. *Pflügers Arch. ges. Physiol.* **371**, 263–269 (1977).
- Cannessa, M., Adragna, N., Connolly, T., Solomon, H. & Tosteson, D. *Clin. Res.* **27**, 511 A (1979).
- Blaustein, M. *Am. J. Physiol.* **232**, c165–c173 (1977).
- De Champlain, J., Krakoff, L. & Axelrod, J. *Circulation Res.* **23**, 479–491 (1968).

Mechanism of action of dietary fibre in the human colon

Alison M. Stephen & John H. Cummings

Dunn Clinical Nutrition Centre, Addenbrookes Hospital, Trumpington Street, Cambridge CB2 1QE, UK

Fibre, more than any other dietary component, affects human large bowel function, causing an increase in stool output, dilution of colonic contents, a faster rate of passage through the gut and changes in the colonic metabolism of minerals, nitrogen and bile acids^{1–6}. (Fibre here refers to 'dietary fibre', which comprises plant cell wall polysaccharides and lignin, and not to 'crude fibre'.) It is thought that these changes are brought about by fibre passing through the gut undigested and holding water within its cellular structure^{7,8}. Although the amount of water taken up *in vitro* varies for different types of fibre^{9–10}, this does not correlate in the expected way with the effects these materials have on colonic function⁹. This is because fibre is extensively degraded in the gut¹¹, probably by the colonic microflora^{12,13}. Using a newly developed method modified from ruminant nutrition for isolating bacteria, we have shown that the main component of human faeces is bacteria¹⁴. We show here that the way in which two contrasting types of dietary fibre act in the colon depends on the extent to which they are digested. Cabbage fibre, which is extensively broken down, provides a readily usable substrate for the stimulation of microbial growth, whereas wheat fibre remains largely undigested and retains water in the gut lumen.

Table 1 shows the effect of wheat fibre in six subjects. On the control diet, faecal solids were 55% bacteria, 17% undigested fibre and 24% water-soluble material. With fibre, stool output increased by 102 g per day (127%), of which 19 g was solids and 83 g water. Transit time decreased by 25.1 h and total nitrogen excretion increased from 1.5 to 2.0 g per day. The 19 g per day increase in excretion of solids was accounted for largely by an increase of 11.6 g per day in undigested fibre. Thus, only 36% of the wheat fibre added to the diet was digested in the gut.

If it is assumed that the extra fibre excreted was hydrated to the same extent as the total stool, which was 3.2 g water per g faecal solids, or 76.3% water, this would account for 48.7 g (48%) of the increase in faecal output. The water-holding

capacity of this wheat fibre, when measured *in vitro*, is in excess of 3.2 g water per g material⁹. Similarly, the extra bacteria excreted (2.3 g) would contribute 9.7 g (10%) of the increase at the same hydration. Bacteria are normally about 80% water¹⁵ and, as they comprise the major fraction of faecal solids, are an important water-holding component in the gut. The whole stool was wetter with the added fibre than on the control diet, which suggests that a further 20.5 g (20%) of the increase was water associated with the bacteria and fibre from the control diet. This leaves 23 g (22%) of the increase to be accounted for by water-soluble solids (5.1 g) and water (18.1 g), which may represent an increase in extracellular, or free, water in the stool.

Wheat fibre, therefore, by surviving digestion in the intestine, alters colonic function by holding water and thus increases the bulk of the colonic contents. Transit time is thereby decreased and less water is absorbed from the lumen. Scanning electron microscope studies of bran before and after passage through the gut suggest it retains its cellular structure virtually intact^{16,17}.

The changes seen with cabbage fibre were quite different. Stool output increased by 54.3 g per day (69%), a smaller increase than with wheat fibre but nevertheless highly significant. Transit time fell by 15.8 h and nitrogen excretion increased from 1.5 g to 2.1 g per day. The increase in stool output comprised 8.5 g per day solids and 45.8 g water. Of the solids only 1.6 g was undigested fibre, indicating that on average 92% of the cabbage fibre had been digested. The main increase was in bacterial solids (4.8 g) which, at the hydration of the stool (74.6%), accounts for 18.9 g (35%) of the increase in stool weight. The faeces were again better hydrated on the cabbage fibre diet than during the control, presumably because the increased bacterial mass stimulated faster transit and led to less water absorption from luminal contents by the large bowel mucosa. These changes in hydration account for a further 17.5 g (32%) of the stool output, the extra fibre excreted adds 6.3 g (12%), and the rest, 11.6 g (21%), is associated with the soluble component.

Cabbage fibre, therefore, influences colonic function through its stimulation of microbial growth. Analysis of the bacterial fraction for nitrogen (for three subjects) shows that 0.42 g of the 0.67 g per day (63%) increase in nitrogen excretion is associated with this fraction, whereas with wheat fibre (four subjects) only 0.18 g of the 0.53 g (34%) increase is in the bacteria¹⁸. The rest of the nitrogen is presumably present in the undigested cell-wall material.

Fibre provides a ready source of available carbohydrate to the microflora of the human colon, as it does in the rumen. On the

Table 2 Changes in faecal composition (g per day $\pm$ s.e.m.) and transit time (h) with cabbage fibre

	Control diet	+Cabbage
Mean daily stool weight	88.2 $\pm$ 9.8	142.5 $\pm$ 16.1*
% Moisture	69.4 $\pm$ 1.9	74.6 $\pm$ 2.0†
Total solids	26.1 $\pm$ 1.2	34.6 $\pm$ 5.7*
Total nitrogen	1.5 $\pm$ 0.1	2.1 $\pm$ 0.1†
Transit time	79.5 $\pm$ 10.4	63.7 $\pm$ 8.0‡
Composition of solids (g per day)		
Neutral detergent fibre	4.1 $\pm$ 0.2	5.7 $\pm$ 0.4*
Water soluble	6.0 $\pm$ 0.5	8.7 $\pm$ 0.7†
Bacteria	14.5 $\pm$ 1.1	19.3 $\pm$ 0.8†

Six healthy male subjects, three of whom also took the bran, undertook the same protocol as outlined in Table 1. Instead of wheat fibre, 18.3 g per day of cabbage fibre was added to the diet.

* $P < 0.001$; † $P < 0.01$; ‡ $P < 0.05$.

basis of calculations from rumen physiology, the digestion of 16.4 g per day of fibre from cabbage would be expected to produce 0.44 g per day bacterial nitrogen, which approximates closely to the findings in this study (0.42 g per day).

Two distinct mechanisms thus emerge whereby fibre affects the human colon. We believe that the stimulation of microbial growth is the more usual one in man because very little fibre survives digestion by the bacteria when sources such as apple, carrot, guar (unpublished observations), pectin¹¹ or mixed diets^{2,21} are fed. Wheat fibre, and possibly cereals in general, may prove to be the exception as they have small cells with highly lignified cell walls which resist digestion.

The increase in excretion of bile acids, neutral sterols, nitrogen, fat, minerals and electrolytes which have been noted in other studies when fibre is fed, have been ascribed to the physical effect of the fibre itself^{2,3,19-23}. Because little fibre is excreted in the faeces, other explanations for these changes must be sought, perhaps taking account of the increase in bacterial mass.

On the control diet, both stool output and transit through the gut were significantly correlated in this study with faecal bacterial mass ($r = 0.84$ and -0.80 respectively). If, as has been suggested²⁴, colonic disease is more common in people who have small stool outputs and slow transit, the control of colonic microflora will be important in determining disease susceptibility in individuals. The type, amount and digestibility of fibre in the diet will, in turn, make a significant contribution to this.

Received 21 November 1979; accepted 25 January 1980.

Table 1 Changes in faecal composition (g per day $\pm$ s.e.m.) and gut transit time (h) with wheat fibre

	Control diet	+ Wheat fibre
Mean daily stool weight	95.5 $\pm$ 8.7	197.0 $\pm$ 13.4*
% Moisture	71.1 $\pm$ 1.5	76.3 $\pm$ 1.2†
Total solids	27.0 $\pm$ 1.2	46.0 $\pm$ 1.1*
Total nitrogen	1.5 $\pm$ 0.1	2.0 $\pm$ 0.1†
Transit time	67.8 $\pm$ 10.1	42.7 $\pm$ 2.7‡
Composition of solids (g per day)		
Neutral detergent fibre	4.6 $\pm$ 0.4	16.2 $\pm$ 0.3†
Water soluble	6.5 $\pm$ 0.5	9.2 $\pm$ 0.4*
Bacteria	15.0 $\pm$ 0.7	17.3 $\pm$ 1.0‡

Six healthy male subjects, aged 22–38 yr, were given a controlled diet for 3 weeks consisting of everyday foods of overall daily composition: 385 g carbohydrate, 85 g protein, 108 g fat and 22 g dietary fibre. For a further 3-week period 18 g dietary fibre as bran was added to the diet. Further details of the study protocol and methods of preparation of the fibre are recorded elsewhere¹. Transit was measured by the continuous marker method²⁵. Faeces were collected throughout, and for each subject those from the last 7 days of each dietary period were pooled and freeze-dried. Daily excretion of fibre, as neutral detergent fibre²⁶, nitrogen by the Kjeldahl technique, bacterial mass and the water-soluble component of faeces were measured¹⁴.

* $P < 0.001$; † $P < 0.01$; ‡ $P < 0.05$.

- Cummings, J. H. *et al.* *Lancet* **i**, 5–8 (1978).
- Southgate, D. A. T. *et al.* *Metabolism* **25**, 1129–1135 (1976).
- Cummings, J. H., Hill, M. J., Jenkins, D. J. A., Pearson, J. R. & Wiggins, H. S. *Am. J. clin. Nutr.* **29**, 1468–1473 (1976).
- Cummings, J. H. *et al.* *Am. J. clin. Nutr.* **32**, 2086–2093 (1979).
- Cummings, J. H., Hill, M. J., Bone, E. S., Branch, W. J. & Jenkins, D. J. A. *Am. J. clin. Nutr.* **32**, 2094–2101 (1979).
- Eastwood, M. A., Kirkpatrick, J. R., Mitchell, W. D., Bone, A. & Hamilton, T. *Br. med. J.* **IV**, 392–394 (1973).
- Eastwood, M. A. & Mitchell, W. D. in *Fiber in Human Nutrition* (eds Spiller, G. A. & Amen, R. J.) 109–130 (Plenum, New York, 1976).
- McConnell, A. A., Eastwood, M. A. & Mitchell, W. D. *J. Sci. Fd Agric.* **25**, 1457–1464 (1974).
- Stephen, A. M. & Cummings, J. H. *Gut* **20**, 722–729 (1979).
- Heller, S. N. thesis, Cornell Univ. (1978).
- Cummings, J. H. *et al.* *Br. J. Nutr.* **41**, 477–485 (1979).
- Bryant, M. P. *Am. J. clin. Nutr.* **31**, S113–S115 (1978).
- Salyers, A. A., Vercellotti, J. R., West, S. E. H. & Wilkins, T. D. *Appl. envir. Microbiol.* **33**, 319–322 (1977).
- Stephen, A. M. & Cummings, J. H. *J. med. Microbiol.* (in the press).
- Luria, S. E. in *The Bacteria* (eds Gunsalus, I. C. & Stanier, R. Y.) 1–34 (Academic, New York, 1960).
- Dintzis, F. R., McBrien, J. B., Baker, F. L. & Inglett, C. E. in *Dietary Fibers: Chemistry and Nutrition* (eds Inglett, G. E. & Falkehaug, S. I.) 157–171 (Academic, New York, 1979).
- Williams, A. E., Eastwood, M. A. & Cregeen, R. *Scanning Electron Microsc.* **II**, 707–712 (1978).
- Stephen, A. M. & Cummings, J. H. *Proc. Nutr. Soc.* **38**, 141A (1979).
- Southgate, D. A. T. & Durnin, J. V. G. *Am. J. Nutr.* **24**, 517–535 (1970).
- Kay, R. M. & Truswell, A. S. *Br. J. Nutr.* **37**, 227–235 (1977).
- Cummings, J. H. *Topics Gastroenterol.* **6**, 49–62 (1978).
- Kritchevsky, D. & Story, J. A. *J. Nutr.* **104**, 458–462 (1974).
- Eastwood, M. A., Anderson, R., Mitchell, W. D., Robertson, J. & Pocock, S. J. *Nutr.* **106**, 1429–1432 (1976).
- Burkitt, D. P., Walker, A. R. P. & Painter, N. S. *Lancet* **ii**, 1408–1412 (1972).
- Cummings, J. H., Jenkins, D. J. A. & Wiggins, H. S. *Gut* **17**, 210–218 (1976).
- Goering, H. K. & Van Soest, P. J. *Agric. Handb. Forest Serv.* **379**, 1–20 (1970).

BOOK REVIEWS

Phylogenetic hypotheses

R. D. Martin

THIRTEEN years have elapsed since the late Willi Hennig's influential textbook *Phylogenetic Systematics*, translated from the German manuscript by Dwight Davis and Zangerl, was originally published in 1966. Republication, with the addition of a brief but useful introduction by three disciples, comes at a time when Hennig's major contribution to the science of phylogenetic reconstruction has more-or-less achieved the widespread recognition that it deserves. This is also a particularly suitable time to take stock of Hennig's approach. Although *Phylogenetic Systematics* is set out with admirable clarity, considerable controversy has accompanied the dissemination and practical application of Hennig's concepts. The issues at stake can only be clearly appreciated by taking a step omitted by Hennig and many others in the field — namely, that of considering separately phylogenetic reconstruction and classification. For the first, Hennig has unquestionably introduced greater scientific precision, while for the second he has arguably generated further confusion.

The attempt to reconstruct phylogenetic history is, or should be, a scientific undertaking involving the identification of past evolutionary relationships ultimately embodied in the process of speciation. One of Hennig's most important innovations here lies in his explicit statement that mere assessment of degrees of morphological (or

Phylogenetic Systematics. By W. Hennig. Pp. 263. (University of Illinois Press: London, 1979.) £12.

other) similarity does *not* provide an adequate basis for accurate phylogenetic reconstruction. Any given group of organisms can be theoretically traced to an original stem species with a certain array of initial (plesiomorphous) characters, and retention of any of these characters as shared similarities among descendant species provides no information about subsequent branching in the evolutionary tree. Only derived (apomorphous) characters developed at some later stage and retained in certain descendants can be used as indicators of phylogenetic relationships within the group. Hennig provides a number of empirical guidelines for the differential assessment of shared similarities (plesiomorphous; apomorphous; convergent) in phylogenetic reconstruction, and concepts such as those of the character transformation series and of the sister group are of considerable heuristic value. However, as Hennig himself recognises, everything in the end boils down to a scientific estimation of probabilities. Thus, the accuracy of phylogenetic reconstruction depends on the validity of the theoretical principles used (hence the value of Hennig's contribution) and upon the diversity of the

characters examined.

Unfortunately, Hennig simultaneously maintains that classifications should be directly based upon the evolutionary branching pattern and the dates of divergence inferred. This controversial tenet has proved to have at least two great drawbacks. Firstly, there is no compelling reason to assume that classifications should only reflect inferences regarding the positions and timing of branching points in phylogenetic trees. Secondly, since every phylogenetic reconstruction represents a hypothesis based on estimation of probabilities, Hennig's prescription inevitably leads to a proliferation of alternative classifications and instability of higher taxonomic categories. Since classifications provide, among other things, the terms we use to discuss living organisms, taxonomic instability is associated with linguistic instability. This being the case, any recommendation to read Hennig's excellent logical dissection of the process of phylogenetic reconstruction must be tempered with the warning that classification (which inevitably involves numerous arbitrary decisions) will not necessarily benefit from too close an association with the phylogenetic hypotheses of individuals.

R.D. Martin is Reader in Physical Anthropology at University College, and Visiting Professor of Zoology at Birkbeck College, London, UK.

For the petroleum geologist

R.C. Selley

Petroleum Geochemistry and Geology. By J.M. Hunt. Pp.617. (Freeman: San Francisco and London, 1979.) £16.95.

THE application of chemistry to the study of oil generation, migration and maturation has a long history. It is only within the last decade, however, that geochemical exploration techniques have become an accepted tool in the oil industry. Dr Hunt has played a seminal role in these advances both in industry, with Exxon,

and latterly at Woods Hole. This book thus merits the attention of every petroleum geologist. It consists of twelve chapters, arranged in four parts: Introduction, Origin and Migration, Habitat, and Applications.

The Introduction begins with a short historical review and continues with an account of the early formation of life on Earth, photosynthesis, the carbon cycle and the preservation of organic carbon in sedimentary rocks. This is followed by a chapter on the chemistry of hydrocarbons, both those which occur naturally and those produced by refineries.

Part II, Origin and Migration, begins with a cursory review of the abiogenic theories of hydrocarbon formation before launching into an account of the chemistry of plant and animal tissues and the

diagenetic changes which they undergo on burial. The temperature required for oil generation is discussed from both theoretical and observational points of view. Gas generation is treated in a separate chapter. The final chapter of this section deals with the last great mystery of oil, namely the magic of primary migration from the source rock to the carrier beds.

Part III, Habitat, opens with an account of source rocks, discussing the oil and gas generating potential of the various types of kerogen, and reviewing the various palaeothermometers available from maturation studies. This is followed by a chapter on the variation and distribution of hydrocarbons in reservoirs, with respect to depth, age and degradation.

Part IV, Applications, shows how the data and concepts previously presented

may be used in hydrocarbon exploration. This begins with a review of surface manifestations of oil, both the visible seeps of oil and gas onshore and offshore, and surface geochemical prospecting. A chapter on subsurface prospecting follows, describing the analysis and interpretation of cuttings gas as well as ways in which source rocks may be studied to determine their oil or gas potential, and their degree of maturity.

The next chapter shows how crude oils may be correlated with their source beds, and the text concludes with a short account of geochemical prospect evaluation. Finally, there is a glossary, forty pages of references and author and subject indexes.

Inevitably, this book invites comparison with Tissot and Welte's recent book, *Petroleum Formation and Occurrence* (Springer: Berlin, Heidelberg and New York, 1978), on the same topic. They both cover the same subject matter and both draw freely (and with due acknowledgement) on each other's work. Both books are excellent and the only simple comparison that may be drawn is that Hunt's

book perhaps emphasises geology more than chemistry, while for Tissot and Welte's book the converse is true. Certainly, this reviewer, who is only a geologist, finds Hunt's book the easier to read.

The arrangement of the vast amount of material in a book of this kind presents many problems but could perhaps have been improved. For example, discussion of source rocks should perhaps precede an account of the emigration of hydrocarbons from them. Sections dealing with palaeotemperature analysis could perhaps have been grouped together, and their relative merits and limitations contrasted, instead of their being distributed through several chapters of the book. Since oil and gas formation is so intimately related it seems an unnecessary distinction to discuss their genesis in separate chapters.

To counteract these comments, however, it must be stated that the book is well written, well researched and pleasantly produced. All the line illustrations have been specially prepared for this work and

each part begins with a half-tone plate. A particularly helpful feature is the inclusion of a summary and list of supplementary reading at the end of each chapter.

The wide range of case histories discussed is impressive. It is unusual to find a geologist of the USA who is familiar with, or at least prepared to quote, data from outside North America. This reviewer could spot only two important omissions. There was no reference to Porfir'ev's paper on the inorganic origin of oil (*Bull. Am. Ass. Petrol. Geol.* 58, 3-33, 1974). This is probably the most accessible exposition of the Russian igneous oil theory for readers in the free world. Several papers related to the correlation of vitrinite reflectivity and electron spin resonance are also omitted.

These are, however, minor blemishes on a major textbook. No petroleum geologist should be without a copy of *Petroleum Geochemistry and Geology*. □

R.C. Selley is Reader in Petroleum Geology at Imperial College, University of London, UK.

Plankton research

C.M. Yonge

Zoogeography and Diversity in Plankton. Edited by S. van der Spoel and A.C. Pierrot-Bults. Pp.410. (Edward Arnold: Maidenhead, 1979.) £37.50.

THIS most attractive book is a revelation of the extent and cohesion of plankton research. It appears at about the centenary of the initial work by Hensen which led, at the beginning of the century, to the formulation by the International Council for the Exploration of the Sea of a programme involving the quantitative study of plankton in relation to hydrography on the one hand and fisheries on the other. These first collections and counting of plankton have gradually spread over the face of globe, with whaling investigations in the Southern Ocean and the International Indian Ocean Expedition extending knowledge into these previously little known areas.

Problems of distribution and of speciation, the dominant themes in this book, are vast and perplexing. In contrast to the terrestrial habitat, that of marine plankton is three-dimensional and continuous. There are notable distinctions between neritic (coastal) and oceanic populations, with latitudinal divisions largely temperature based and different biotopes in depth represented by epipelagic, mesopelagic and bathypelagic plankton. Distinctions between habitats are often best defined by the resident 'indicator' species, most frequently

chaetognaths. The *Challenger* studies of Murray and Renard on the deposition of planktonic skeletons on the ocean floor are the basis of modern work on the history of ocean basins and provide important evidence concerning continental drift.

An adequately authoritative survey of this immense field demands specialist treatment and most exacting editorial standards. Both are supplied by the contributions of fourteen authors from seven countries, most competently

organised by S. van der Spoel and A.C. Pierrot-Bults of the Institute of Taxonomic Zoology, University of Amsterdam. Clearly drawn figures and an extensive literature list complete the attractions of a book which presents information essential to all interested in the full extent of biological enquiry. It makes immediate appeal to the scientific imagination. □

Sir Maurice Yonge is Honorary Fellow in Zoology at the University of Edinburgh, UK.

Light physics

P. W. Hawkes

Optoelectronic Devices and Optical Imaging Techniques. By D. A. Ross. Pp. 137. (Macmillan: London, 1979.) Hardback £12; paperback £4.95.

THIS clear, simple introduction to a range of devices that convert electrical energy into light or generate such energy from light will appeal to a wide student audience, particularly to those who need to acquire only a smattering of the physics involved before exploring the many applications. The text is orientated towards specific devices, chapters being devoted to LEDs, photoconductors, photodiodes and phototransistors, solar cells and laser diodes; two other chapters are concerned with noise in such devices and in optical recording media, first film and then CCDs. All these are described in easy, uncomplicated language and illustrated

with uncluttered line diagrams and a photograph showing CCD-aided decapitation of carrots! I was a little startled to encounter, for the first time for many years, the jovial, avuncular style of older school textbooks; a line diagram is described as "An artist's impression of a photon", we are told that "It would be a minor disaster to be forced to abandon Fourier analysis . . .", but isolated examples cannot properly convey the flavour which nonetheless pervades the book.

The publishers have sensibly also brought out the book in paperback. Libraries will of course be more interested in the cased edition, but the paperback version is good value at £4.95; it should sell well, not only to students but to others, since plenty of people are interested in understanding these devices by which they are surrounded. □

P.W. Hawkes is Maître de Recherches in the Laboratoire d'Optique Electronique du CNRS, Toulouse, France.

Transfer RNA

J.P. Goddard

Transfer RNA: Structure, Properties, and Recognition. Edited by P.R. Schimmel, D. Söll, and J.N. Abelson. Pp.577. (Cold Spring Harbor Laboratory: New York, 1979.) \$60.

NEARLY a quarter of a century has passed since the existence of the small adapter molecules known as transfer RNA was predicted. Their subsequent study has been motivated by their properties. They are relatively small and their central role in protein synthesis involves many specific interactions with proteins and nucleic acids. Furthermore they have since been strongly implicated in regulation of gene expression. The extent of this study is reflected in the fact that two volumes — the second is on biological aspects — were found necessary to fulfil the aim of summarising the knowledge of tRNA up to about the end of 1978.

The means to that end is, as in previous Cold Spring Harbor monographs, a series of articles by specialists in a particular aspect of tRNA who participated in the Transfer RNA Meeting at Cold Spring Harbor in August 1978. These articles have been supplemented by four appendices — on tRNA sequences and a proposed numbering scheme, on modified nucleosides and their chromatographic mobilities, and the characteristics of aminoacyl-tRNA synthetases. This first volume contains some thirty-one articles covering aspects of primary structure and isolation, crystal structure analysis, solution studies, aminoacyl-tRNA synthetases, synthetase recognition, and interaction with the ribosome.

Unlike several of its predecessors, this monograph contains no general review. My immediate reaction to this omission was to praise the economy of the editors in not adding to the several reviews published in the last couple of years. However, a non-specialist seeking in this volume a thumbnail sketch to tRNA research will be disappointed. A short article expanding the contents of the brief preface by putting each article in its context and perhaps citing other recent reviews would have added less than one per cent to the volume's bulk and price but would have increased its value both to new workers in the field and to scientists from other specialities.

This lack is compensated to a large extent by articles which have extensive bibliographies. This is particularly true of the articles beginning each section which are written by the session chairmen at the meeting who refer to articles later in their section. The invited contributors of articles have usually endeavoured to present not only their own work but also that of other

workers in the field. As the writers are, in the main, reporting topics on which they are among the leading experts, many quite reasonably present reviews of their past few years' work interspersed with that of others. This occasionally leads to a one-sided interpretation in areas where there is currently disagreement, e.g. the structure of the *E. coli* ribosome. However, the volume is generally notable for the generosity and objectiveness of the writers in reporting views and ideas contrary to their own. These virtues are not purchased at the cost of dull uniformity. Each article retains its individuality and yet the volume forms a coherent unit.

In the absence of a moratorium, it is fortunately not possible to produce "the

definitive reference work on transfer RNA". Even the proposed numbering system of nucleosides in tRNAs (Appendix IA), so eminently reasonable a year ago, looks strangely dated in the light of the structure of the tRNAs in our own mitochondria. However, this volume is the best we can hope for — a summary of the major results of past tRNA research with hints of future solutions to outstanding problems. It should prove stimulating reading to all those working in tRNA research, and those working in other aspects of gene expression will find many articles of value. □

J. P. Goddard is a Lecturer in Biochemistry at the University of Glasgow, UK.

Unifying the forces of nature

Malcolm MacCallum

Elie Cartan and Albert Einstein: Letters on Absolute Parallelism, 1929-32. Edited by R. Debever. Translated by J. Ritter and J. Leroy. Pp. 233. (Princeton University Press: Princeton, New Jersey, and Guildford, UK, 1979.) £11.40.

I FOUND these 38 letters (and one postcard) an enjoyable and interesting read, but it is difficult to discern the intended readership, and so I do not know to whom I could recommend the book. The reader needs to be polyglot. Einstein wrote in German, Cartan in French. Both are translated, Ritter presumably translating Einstein's letters and Leroy those of Cartan. The preface and notes by the editor, the well-known Belgian relativist, Robert Debever, are, however, given only in French. The translations are passable, though Leroy's seems to show that English is not his mother tongue. The reader also needs a good knowledge of the background, and the energy to read the references, as the book is not self-contained. The editor even omits one manuscript note because a slightly modified version was published, and he gives only bibliographical details of most of the printed material the correspondents exchanged.

The letters cover three main topics. The first is the writing of a historical note on spaces with absolute parallelism which Cartan drafted in only two weeks in May 1929. It appeared, accompanying Einstein's paper on his attempt at a unified field theory using such spaces, in 1930; Einstein had formulated his ideas in ignorance of the mathematical literature. The second, covered by 26 letters between 3 December 1929 and 17 February 1930, deals with alternative systems of field equations in such spaces, and qualitative features of such systems of partial

differential equations linear in first derivatives. Cartan showed there was another set of equations with the same mathematical properties as Einstein's (involuntary character, determinacy of the Cauchy problem, and arbitrariness of initial conditions), and Einstein gave reasons for thinking them less physically satisfactory. This discussion gradually narrowed to the third subject, the question of whether one could define the 'strength' of a set of such equations by counting freely specifiable functions of r variables for every r up to some maximum. Cartan argued that only the maximal r gave a meaningful result. The two authors never agreed, even in the seven later letters exchanged in 1931 and 1932. By 1932 Einstein had abandoned distant parallelism for the Einstein-Mayer theory, and the only later outcome is the much modified idea of 'strength' used in *The Meaning of Relativity* (fifth edition, Appendix II) which Debever does not cite.

What use, then, is the book? The modern researcher, seeking a unified field theory, will find little of direct relevance to present efforts in this direction. The professional historian of science would probably prefer to consult the original manuscripts. The general reader may be put off by the effort required of him. If he is not, then he may be rewarded in several ways. There are interesting personal sidelights on the characters. The speed of the mail, and the authors' energy (especially as Einstein was 50 and Cartan 60) must be admired. There is much to learn from Einstein's attitude to physical problems and from Cartan's clear and careful mathematical expositions, and some reassurance for lesser mortals in their occasional mistakes. Principally, one can understand how two great minds worked in developing an (albeit unsuccessful) attempt at one of the greatest problems in physics, the unification of the forces of nature. □

Malcolm MacCallum is a Lecturer in Applied Mathematics at Queen Mary College, University of London, UK.

OBITUARY

Sir Barnes Wallis — 1887-1979

I FIRST heard of Wallis in 1920, when the airship R 80, for the design of which he was responsible, began its trials. In retrospect, this was the most successful rigid airship designed in this country but the Air Staff decided that they had no use for it and it flew for the last time on 20 September 1921. As a result of this decision, Wallis was forced to leave the world of design in which he had already, at the age of 34, proved himself a master. He took up teaching in Switzerland and began his courtship of the lady who was to become his wife. In 1923 however a new association with rigid airship design began. Another remarkable man, Commander (and ultimately Sir) Dennistoun Burney, had been at work during Wallis's absence abroad, gaining support for his airship project. The result of Burney's persistence was the creation of the Airship Guarantee Company, a Vickers subsidiary headed by himself with headquarters at Howden in Yorkshire, and the renaissance of the Royal Airship Works at Cardington. Wallis became the chief designer at Howden and at Cardington the design team was led by V.C. Richmond. The Air Ministry gave the Airship Guarantee Company the task of building R 100 and the Royal Airship Works the task of building R 101.

In 1924 work began in earnest on both designs. In outline they were similar, each five times as long as its greatest diameter, with that diameter at 40% of the length from the nose. They were both to hold 5,000,000 cu.ft. of hydrogen. In 1924 there also began an absurd segregation of the two design teams. I was conscious of it almost from my first day at Cardington, in July 1924. That a tremendous competition should begin in that year was inevitable, and indeed desirable, but that each team should find itself with a boundary wall round it, and have no opportunities to compare notes and progress, was absurd and damaging: each team would have benefited by knowledge of the ideas and progress of the other. Where the blame for this situation, which rapidly developed and was sustained to the end — the end of R 101 — should lie I am uncertain. Wallis, already a successful airship designer, highly individualistic, formidable in his belief — not unjustifiable — in his own ideas, was almost the antithesis of Richmond, whose experience had been with non-rigid airships and whose merits were rather those of a manager and a talent-spotter than of an aircraft designer. Each was scornful of the other's efforts. In

Camera Press

my five years at Cardington I do not think they ever met. But the long-range animosity — it was no less — was fostered by others, and if Richmond was no genius as Wallis was, he nevertheless had an *eminence grise*, Flt Lieut. (later Sqdn Leader) F.M. Rope at his elbow whom, had Wallis ever met him, he would have had to admit to his own extraordinary intellectual level.

In R 100 Wallis, I believe for the first time, exhibited his passion for helices. The main booms of the heavy triangular section girders were made by rivetting duralumin strip on itself to produce a tube resembling a paper drinking straw. Wallis designed the special rivetting machine for the job. The wiring conveying the lift of the gasbags to the joints of the structure was a system of helices — or, more accurately, geodetics — in which the mesh became finer as the height from the bottom of the ship, and hence the gas pressure, increased. The affection for the helix, which in Wallis's hands produced elegant solutions to structural problems, manifested itself even more dramatically in his two most famous aeroplanes, the Wellesley and the Wellington.

Before the splendid flight of R 100 across the Atlantic and back in the summer of 1930, Wallis had been persuaded by Sir Robert McLean, the managing director of Vickers, to leave Howden and join the

aircraft design side of Vickers. His first appointment was at the Supermarine works at Southampton where he did not cooperate successfully with another remarkable man, R.J. Mitchell, the designer of the Spitfire. After a few months he was moved to Weybridge to work alongside still another great man, R.K. Pierson, famous for a whole series of aircraft, including the Vimy which made the first transatlantic flight in 1919. They were both ranked as Chief Designer.

The association of Pierson and Wallis was a complete success. Pierson was a big man, in physique and in character: soon Wallis, four years the elder, whom at this time I began really to know, called him Uncle Rex, which I think defined Pierson's relations with the major personality so suddenly thrown into close association with him.

While Wallis was back at Howden during the preparations for R 100's Atlantic venture, he was not allowed by McLean to fly in her, which seems a little hard. But by the time he had left Howden for Southampton, Wallis indeed had already begun to doubt the future of the airship in the face of the dramatic progress of the aeroplane, so perhaps he did not feel the deprivation too deeply. He was soon back at Weybridge thinking about aeroplanes and developing the system of geodetic construction.

IMAGE
UNAVAILABLE FOR
COPYRIGHT
REASONS

At that time I was working on criteria defining the stiffness characteristics of aeroplane wings necessary for avoiding wing flutter and loss of aileron control due to wing twisting. Wallis came to see me and said, in effect, "Tell me how stiff to make my Wellesley wing and I will make it that stiff and no more". I told him and he did indeed achieve the minima I prescribed with a structure which was extremely light. The structure weight he achieved was well below what was being achieved in wings of more conventional construction.

I have never been convinced that Wallis's geodetic system was intrinsically superior to more conventional systems. It was superior in practice because he designed extremely efficiently to the prescribed limits. I have no recollection of any other designer of those days seeking at the outset of his design to discover the stiffness limits to which he should design. Even if I do dare to doubt the intrinsic superiority of geodetic construction however, I certainly do not deny its elegance. It had moreover a tremendous advantage in battle. The fact that strength and stiffness were, so to speak, distributed in what was almost a skin of geodesic mesh meant that the structure had few weak points, and gunfire from the ground and from the air might knock great holes in the structure without wrecking it. So triumphant was it that one might well ask why it did not persist as a basic system for aircraft construction. The answer is that the whole trend of design was away from frameworks covered with fabric and towards all metal monocoque construction. That is why the Wellesley and the Wellington were never emulated, even by Wallis himself.

The Wallis I knew was never the arrogant

prickly person we of the R 101 design team were led to visualise. Certainly his was a strong character, certainly he was firm in his views, but equally certainly he knew his limits and did not hesitate to seek help in areas beyond them — as when he sought my aid on wing stiffness — and I am sure he sought Pierson's aid, particularly on the aerodynamic side of his designs. He was indeed sometimes critical of others working in his own areas of expertise, but who is not? I always found him supremely articulate, subtly persuasive and invariably polite, all in a voice it was a pleasure to listen to. I found him, too, appreciative of the work of others. I recall that it was with some trepidation that I took him, at his request, to Hucclecote to see the drawings of the Gloster E 28/39 which Carter was designing for the first flight of the Whittle jet engine. I thought there might be questions of the "Why didn't you do it this way?" kind. There were not. Carter was completely open — he displayed all his ideas and Wallis had nothing but commendation then and afterwards.

I had no contact with Wallis on the Mohne and Eder dambusting bomb nor its massive successor. But I was in touch, as an occasional visitor to Weybridge, with his ideas on the swing wing or variable geometry aeroplane and bitterly disappointed at the lack of support for his ideas. The swing wing has been with us for some years now, but had it been adopted in practice when Wallis showed the way this country would have had a commanding technological lead in a new field: once again, we threw away a golden opportunity.

In his office on the old Brooklands track, and after his retirement from

Vickers in his home, he always worked out his ideas himself on his drawing board. Obviously, on his great projects, he directed the work of others. But the intricacies of joints in geodetic meshes, the details of the pivot of a swing wing, were elucidated on his own board. It was the same to the end: when I last saw him in his house at Effingham he was working on his board on very high speed vehicles indeed. I can only hope his efforts to bring these last visions to reality will not be wasted.

I have made no attempt in these notes to chart the long life of Barnes Neville Wallis, from his birth on 26 September 1887, to his death on 30 October 1979. The fullest account, up till 1970, is Mr Morpurgo's biography¹. I do not agree with quite all of Mr Morpurgo's judgements — nor indeed with quite all of those he attributes to Wallis — but factually it is a vivid and accurate account of a great man's life and I hope the biographer will add a chapter to complete the story. In that book we can read the true account of how Wallis was awarded by the Royal Commission on Awards to Inventors £10,000 for his dambusting bomb, and how he applied that sum to the creation of a foundation at his old school, Christ's Hospital — of which he was Treasurer from 1957 to 1970. This splendid act illumines facets of Wallis's character with which the world in which he became famous was unfamiliar — his loyalty, his generosity, indeed his unselfishness. In total he was a great and good man. We who knew him were greatly privileged.

Kings Norton

1. Morpurgo, J.E. *Barnes Wallis: A Biography*. (Longman, London 1972).

Wolfgang Yourgrau

WOLFGANG H.J. YOURGRAU died suddenly at his home in Denver, Colorado, USA, on 18 July 1979, at the age of 70, shortly after returning home from an extended tour of lecturing and research in Europe. His passing marks the disappearance from the world of physics and the philosophy of science of a colourful scholar and humanist, whose diverse talents and activities have left enduring imprints far outside the field of his speciality.

Wolfgang Yourgrau was born on 16 November 1908 near Berlin of a Belgian father and a German-Jewish mother. He attended the Werner-Siemens Realgymnasium in Berlin and subsequently studied theoretical physics, mathematics, and biology at the local von Humboldt University, at a time when celebrated physicists such as von Laue, Einstein, and Schrödinger graced its Faculty. Serving first as a tutor in natural philosophy and then as an assistant to Schrödinger, he earned his Dr phil. *magna cum laude* in

1932, the terminal year of the Weimar Republic.

The next year saw the ascendancy to power of the Nazis, and Yourgrau, who got their attention as an organizer of the S.A.P. (an offshoot of the German Social Democratic Party), fled Germany after being severely beaten up by the Storm Troopers. In exile, he remained on the move while lecturing on the evils of fascism in Latvia, Poland, and other European countries, until he was allowed to enter Palestine, then a British mandate, as a political refugee.

Appointed a lecturer with the educational and cultural division of the *Histadrut* (Jewish Federation of Labor), he travelled widely inside Palestine and became drawn into discussions on the intractable problems of this territory. But Nazism, its devastation of the cultural values in which his family had been rooted for centuries, and then the daily progress of World War 2 remained foremost in his mind. In the Spring of 1942 he acquired from the British authorities a license to publish *Orient*, an independent German-

language weekly, with himself as Editor-in-Chief and Arnold Zweig, the renowned novelist-dramatist and a fellow exile, as co-publisher. Under Yourgrau's direction, *Orient* — which has lately become the subject of detailed analyses by some literary historians — declared war "on every fascist movement, every attempt to restrict the right to free expression of opinion . . ." But it was a subsidiary point on the programme of *Orient*, touching on the complex internal politics of Palestine, which made the journal a centre of intense controversy. Boycotts and threats against businesses connected with *Orient* and finally destruction of the premises of its fourth printer caused the financially plagued journal to close down for good in 1943, one year after its inception.

Yourgrau lost no time in assuming a more active role in the war against the Axis Powers, which was then increasingly being fought in the Mediterranean theatre. At first, working from an office of the Palestine Information Office, this only meant preparing the news bulletins in

German that emanated daily from Jerusalem Radio. But after the American Office of Strategic Services established its Middle Eastern Headquarters in Cairo, Yourgrau also was recruited to serve with that organization, vitally assisting it in its planning of some intelligence operations behind enemy lines, details of which are only now being declassified.

At War's end Yourgrau redirected his phenomenal energy into the resumption of his academic career and the pursuit of his pre-empted research. He became Head of the Department of Logic and Scientific Method at the School of Higher Studies in Jerusalem and subsequently Acting Dean of its Faculty of Arts and Sciences. At the request of the Colonial Office he went to Cyprus in 1948 to determine whether a branch of the University of London should be opened on that island. ("Under no circumstances!" his report concluded after 4 months of investigation.) In the same year Yourgrau emigrated to South Africa, having previously married South African-born Thella Garber. For a decade he taught and continued his writing and research at the Universities of Cape Town, Witwatersrand, and Natal.

In 1959 he moved to the United States, first to accept the position of Research Professor at the Minnesota Centre for the Philosophy of Science, and then to become Chairman of the Department of History of Science at Smith College in Northampton, Mass. In 1963 he accepted a permanent position as Professor of History and Philosophy of Science at the University of Denver.

The scope of Yourgrau's publications was prodigious. Ranging from the political editorials in *Orient* to papers on general relativity, they made him known to an exceptionally large spectrum of scholars, a fact that is attested to by the scheduled appearance in 1981 of a memorial volume of essays in his honour written by a diverse and distinguished panel of more than 30 of his academic colleagues. Although most of his papers are devoted to problems in theoretical physics, a large fraction deals with philosophical issues, while some others treat matters of a biographical, literary, or political nature. Of the many books he co-authored or co-edited, *Variational Principles in Dynamics and Quantum Theory* and *A Treatise on Irreversible and Statistical Thermodynamics* are perhaps the best known. In 1969 he founded, with Henry Margenau of Yale University, the international journal *Foundations of Physics*, which he co-edited until his death. He was the recipient of numerous distinctions and honours, which included the Swiss Einstein Medal, awarded to him in 1970.

Gregarious and extrovert, his forthrightness and lack of false modesty endeared him to some, made enemies of others. His co-workers and many friends in different parts of the world enjoyed his unique kind of humour, were buoyed by

his passion for life, and stimulated by his enthusiasm for intellectual pursuits. Sentimental and deeply emotional, he was intensely loyal to individuals whose friendship he valued, expecting the same degree of allegiance in return. But perhaps the most enduring impression will be his automatic reflex to side with, and concretely support, human beings — whether fellow students in Germany, penniless intellectual exiles in Palestine, or anyone else who crossed his path — who were treated unjustly or were in need of help. He is survived by his wife, a daughter, and three sons, to whom we extend our profound sympathy in their great personal loss.

Alwyn van der Merwe

David Scott Gilbert

DAVID SCOTT GILBERT was born in Ithaca (N.Y.) on 6 November 1940. He majored in mathematics at Harvard University (1959-63) then, perhaps because of a biological tradition in his family, joined D.H. Fender at the California Institute of Technology, where he gained his doctorate on visual acuity and eye movements. His postdoctoral work was with the late Trevor Shaw at Queen Mary College London, where he later became a lecturer in zoology before joining, in 1973, the Medical Research Council Cell Biophysics Unit at King's College London. He died suddenly of viral pneumonitis on 11 December 1979.

While working with Shaw on sodium transport across the membranes of giant axons, Gilbert became interested in the structural properties of axoplasm and thus concerned with problems of the determination and maintenance of the shape of nerve cells. His first paper on axoplasmic structure in 1972, (*Nature, New Biology* 237, 195-198) was remarkably stimulating and procedurally elegant. Using polarised light microscopy he showed that the giant axon of *Myxicola* (a marine fan worm) can be described in terms of three levels of helical organisation. The axoplasmic fibrous proteins are arrayed in parallel 'ripple' helices, which are twisted into a larger 'segmental' helix. They form a cylindrical gel that can coil, when the worm shortens during contraction, to form a yet larger 'gel' helix. Gilbert recognised that this axon consisted of essentially one structural component, the neurofilament, and therefore it provided an unparalleled opportunity for experimental study. The demonstration of the helical organisation led to a model in which the filamentous protein forms into twisted strands like a rope and, although the protein content of the axoplasm of *Myxicola* is no more than about 4%, the yarn gives it significant

mechanical strength and stability; and it can shear to form branches.

Many earlier observations of axons had been interpreted as demonstrating them to be a viscous soup, rather than this kind of stiff gel, and had led to the supposition that the mechanical properties were due to the membrane and associated connective tissue. Gilbert and his colleagues found the neurofilament gel to be solubilised by a very rapid enzymic autolysis triggered upon the entry of calcium ions into the axoplasm; the technique of very quickly extricating the axoplasm in air had avoided exposure to the calcium ions of physiological saline or sea water. This method enabled Gilbert (1975) unequivocally to determine many of its bulk chemical and physical properties (*J. Physiol. Lond.* 253, 257-319). Thus the preparation provided an important standard of reference for working on the biochemistry and structure of filaments (10nm filaments, *Nature* 272, 557-8) and formed a firm basis from which a long-needed attack on axonal structure at the molecular level could begin.

Gilbert was leading the developments on a wide experimental front at the time of his death. Fortunately much of his work with that of colleagues is drafted for publication. Using mammalian as well as cephalopod and annelid nerve cells they have uncovered a wide variety of different neurofilament proteins. To examine their homologies they have developed a new high resolution system of gel electrophoresis and are beginning to show by fingerprinting techniques that the variety of neurofilament chains in several species, including some mammals, conceals an underlying simplicity. With a view to uncovering the biochemical as well as the structural features of neurofilaments Gilbert and his colleagues have assayed their modification by the action of axonal proteases and of endogenous kinases and phosphatases. At the same time data from X-ray and solution studies are leading to the development of a detailed molecular model of the neurofilament.

Gilbert had a keen critical intelligence and an uncommon ability to choose significant and clear-cut biological problems. He mastered a wide variety of techniques and brought to them exceptional manipulative skill. In our laboratory his broad interests and strong grasp of physical principles made him much in demand for discussion and advice. He was infallibly helpful and generous. We, his colleagues, are impoverished by his death, which will grievously set back the development of this kind of neurobiology in Britain. David had an essential humility and natural friendliness that enabled him to enjoy an easy relationship with a wide variety of people. We mourn him as much because we have lost a friend, as because a young scientist of distinction has vanished from the international scene.

B.B. Boycott